ROLE OF AWARENESS AND METACOGNITION IN AUDIO-VISUAL VENTRILOQUISM

by

PATRYCJA SYLWIA DELONG

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
September 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This unpublished thesis/dissertation is copyright of the author and/or third parties.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

Human perceptual experience is inherently multisensory. The brain continuously makes decisions about which sensory signals should be interpreted as common and which as separate events. Basic principles of the processes underlying such decisions have been widely explored over the past few decades. While there is a good understanding of the function of bottom-up binding cues, little is known about how automatic multisensory integration is and what is the role of subjective perceptual experience. This thesis employs the Ventriloquist Effect (VE) to investigate the interplay between awareness, metacognition and audio-visual integration. Chapters 1 and 2 describe the current state of knowledge about relevant aspects of multisensory integration and analytical techniques applied in the empirical chapters. Chapter 3 shows that Ventriloquist Effect can be caused by visual stimuli presented under Continuous Flash Suppression, which are not consciously perceived. Chapter 4 investigates neural mechanisms of the effect for aware and unaware flashes. Chapter 5 shows semantic modulation of VE for unmasked, but not masked images, even though VE is present for unseen stimuli. Chapter 6 investigates the function of spatial and semantic congruency in multisensory causal metacognition. Chapter 7 provides a summary of the main findings and discusses their contribution to the field.

AKNOWLEDGEMENTS

I would like to thank to all the people that helped me survive this challenging time in my life, that I spent working on this PhD:

- my supervisor Uta Noppeney for lengthy discussions, guidance and all the helpful advice regarding experimental designs, analytical methods and academic writing.
- all my colleagues from the Computational Cognitive Neuroimaging Lab for great conversations (about more or less scientific topics) and sharing ups and downs of our lives as PhD students (in no particular order): Arianna Zuanazzi, Sam Jones, Remi Gau, Steffen Burgers, Ambra Ferrari, Agoston Mihalik, Mate Aller, David Meijer, Giulio Degano, Martin Šmíra, Alex Murphy and Michael Joannou.
- my partner Guy for proofreading this thesis, adding all those missing articles everywhere, and for being the most supportive and understanding during the final weeks of thesis writing – without him I'd probably live on frozen pizza.
- all volunteers that completed my lengthy and tedious EEG experiments – your effort is much appreciated!
- my family for ongoing support irrespective of what I wanted to do at various points of my life.

Table of Contents

List of Figures:

Chapter 2

Chapter 3

Chapter 4

Chapter 5

Chapter 6

List of Tables:

Chapter 2

Chapter 4

Chapter 5

Chapter 6

List of Abbreviations:

VE	Ventriloquist Effect
BCI	Bayesian Causal Inference
SOA	Stimulus Onset Asynchrony
TWIN model	Time window of integration model
ERP	Event related potential
MMN	Mismatch negativity
fMRI	functional Magnetic Resonance Imaging
EEG	Electroencephalography
BOLD	Blood Oxygen Level Dependent contrast
CFS	Continuous Flash Suppression
dCFS	dynamic Continuous Flash Suppression
GWT	Global Workspace Theory
CMB	Crossmodal bias
SDT	Signal Detection Theory
BCI	Brain-Computer Interface
PET	positron emission tomography
MVPA	Multivariate pattern analysis
knn	k-nearest neighbours
LDA	linear discriminant analysis
SVM	support vector machine
SVR	support vector regression
GNB	Gaussian Naïve Bayes
MCC	multiple comparison correction
TFCE	threshold free cluster enhancement
GNW	Global Neuronal Workspace
PAS	Perceptual Awareness Scale
ICA	independent component analysis
ANOVA	analysis of variance

CHAPTER 1. INTRODUCTION

Human perceptual experience is inherently multisensory. We constantly receive information from multiple sensory inputs, which are utilized to interact with our environment. Within this cacophony of signals, the brain constantly has to make decisions, which of those signals go together and which are caused by separate events. Whether we are crossing a road, looking for a missing cat in the park or trying to understand a friend talking in a noisy pub, we rely on and integrate information from multiple senses.

Combining sensory inputs coming from common sources allows us to process information about the environment more effectively. Multisensory integration can significantly improve reaction times and performance in perceptual tasks - a phenomenon called multisensory enhancement (Arnold, Tear, Schindel, & Roseboom, 2010; Diederich & Colonius, 2004; Freeman, Wood, & Bizley, 2018; Hairston, Laurienti, Mishra, Burdette, & Wallace, 2003; Lovelace, Stein, & Wallace, 2003; Nidiffer, Stevenson, Krueger Fister, Barnett, & Wallace, 2016). The ability to effectively bind multimodal signals can be crucial for survival and integration mechanisms seem to work similarly across the animal kingdom. For example, multisensory enhancement was shown not only in humans, but also in nonhuman primates (Frens & Van Opstal, 1998), ferrets (Hammond-Kenny, Bajo, King, & Nodal, 2017) and even in chickens (Verhaal & Luksch, 2016).

Not every multimodal presentation results in the integration of signals. Whether any two, or more, sensory signals are perceived as caused by one or more events depends on a variety of low and high order combination cues. In this thesis the ventriloquist illusion was employed to allow clear discrimination between integration and segregation of physically identical stimuli pairs. The Ventriloquist Effect is a perceptual shift in sound localization, towards a concurrently presented visual signal (Jackson, 1953). The illusion is a result of audio-visual binding, therefore sensory integration can be inferred when observers report experiencing ventriloquism.

While binding cues were extensively researched over the years (see review below), very little is known about how automatic multisensory integration is and how it can be influenced by subjective sensory experience i.e. in varying states of perceptual consciousness and confidence. Here, in 4 experimental chapters, I investigate the role of perceptual awareness and metacognition in audio-visual integration. First, the theoretical background is outlaid in this introduction (including: causal inference in multisensory perception, properties of the ventriloquist illusion, the role of awareness and metacognition in multisensory integration), and methodological foundations of applied research techniques are detailed in Chapter 2. The four subsequent experimental chapters describe conducted psychophysics and EEG studies. In Chapter 3, I ask whether the ventriloquist effect can be elicited by unaware visual stimuli (for subjective and objective awareness criteria). In Chapter 4, I explore the neural mechanisms of ventriloquism for visible and invisible visual signals. Next, in Chapter 5, I investigate to what extent signals from different sensory modalities can interact in the absence of perceptual awareness and test whether semantic congruency of audio-visual signals modulate the strength of the ventriloquist illusion. In Chapter 6, I examine the influence of spatial and semantic audio-visual correspondences on multisensory causal metacognition. Finally, in Chapter 7, I summarize the results and discuss their importance.

Causal inference in multisensory perception

In the dynamic environment, the brain constantly needs to make decisions whether or not to bind information from different sensory modalities and determine which signals belong to the same object. Most prominent models of multisensory integration include Maximum Likelihood Estimation and Bayesian Causal Inference (Ernst & Bülthoff, 2004; Shams & Beierholm, 2010). Maximum Likelihood Estimation, also called a forced fusion model, can describe sensory integration reasonably well (Alais & Burr, 2004; Ernst & Banks, 2002), but the Bayesian Causal Inference model emerges as more efficient in predicting behaviour (Beierholm, Körding, Shams, & Ma, 2008; Körding et al., 2007; Meijer, Veselič, Calafiore, & Noppeney, 2019). The BCI model describes the entire process of causal decision making, acknowledging that a pair of stimuli can be either processed together or separately. Factors that determine perceptual binding of two senses depend on integration priors (i.e. observer's expectations of signals coming from a common source) and the signals' properties such as their temporal coincidence, spatial discrepancy and crossmodal correspondences.

When the stimuli are integrated, the brain computes sensory estimates of the unified percept using unimodal estimates weighted according to their sensory reliabilities (e.g. Alais & Burr, 2004; Charbonneau, Véronneau, Boudrias-Fournier, Lepore, & Collignon, 2013; Rohe & Noppeney, 2015b; Sheppard, Raposo, & Churchland, 2013). For instance, spatial estimates of audio-visual stimuli depend on relative sensory uncertainties and perceived location is shifted towards the more reliable stimulus. Vision provides more reliable spatial estimates than audition, which results in the ventriloquist illusion (Alais & Burr, 2004). The illusion will be less prevalent if visual sensory noise is increased (Rohe & Noppeney, 2015b), and a shift towards auditory stimulus can be observed if reliability of auditory estimates is higher than visual (Alais

& Burr, 2004; Bertelson & Radeau, 1981). Another example of auditory bias on vision that results from reliability weighted integration is sound-induced flash illusion. In temporal numerosity judgement, we observe lower variance in audition than vision, and when observers are presented with a single flash and multiple beeps, they report seeing multiple flashes (Shams, Kamitani, & Shimojo, 2000; Shams, Ma, & Beierholm, 2005). Similarly in the time domain, for temporal judgement tasks, the time the visual stimulus is perceived shifts towards the time when an asynchronous auditory distractor is presented – so-called "temporal ventriloquism" phenomenon (Bertelson & Aschersleben, 2003; Morein-Zamir, Soto-Faraco, & Kingstone, 2003). Auditory dominance has also been shown in the temporal bisection task, where interval duration judgment was influenced by the time of sound presentation (Burr, Banks, & Morrone, 2009).

Prior knowledge and common source expectations (which influence the integration prior) can facilitate multisensory integration (Gau & Noppeney, 2016; Helbig & Ernst, 2007; Sarmiento, Shore, Milliken, & Sanabria, 2012; Van Wanrooij, Bremen, & John Van Opstal, 2010). For example, if bimodal stimuli are presented together more often in some experimental runs (congruent context) than the other (incongruent context), common source expectations should be higher in the first case. Stronger crossmodal effects in the congruent context were shown for duration judgement task, where perceived duration of visual stimulus presentation was shifted towards duration of simultaneously presented sound (Sarmiento et al., 2012). A decrease in audio-visual binding was observed for the McGurk effect after incongruent audio-visual speech presentation (Nahorna, Berthommier, & Schwartz, 2012, 2015). The McGurk illusion is a phenomenon, where speech perception is altered by concurrently presented video showing conflicting lip movements (i.e. lip movements saying "ga", can change perception of auditory “ba” into “da”) (McGurk & MacDonald, 1976). Van Wanrooij and colleagues manipulated the probability of spatially congruent and incongruent audio-visual stimuli presentations. Participants responded faster in blocks in which presented stimulus pairs were more likely to be spatially aligned (Van Wanrooij et al., 2010). Similar modulation of expectations in an experiment using McGurk stimuli presented in congruent and incongruent blocks (with more or fewer phonologically congruent trials in a block respectively), showed an increase of audio-visual integration (i.e. more McGurk percepts) in congruent comparing to incongruent contexts (Gau & Noppeney, 2016).

Spatial and temporal congruency are crucial when it comes to determine whether two signals had one common or two separate sources. The percentage of integrated stimuli (e.g. reported as having a common cause) gradually decreases with increasing stimulus onset asynchrony and/or distance (Lewald, Ehrenstein, & Guski, 2001; Lewald & Guski, 2003, 2004; Munhall, Gribble, Sacco, & Ward, 1996; Stevenson & Wallace, 2013; van Wassenhove, Grant, & Poeppel, 2007; Wallace et al., 2004). Multisensory gain also declines with increasing temporal asynchrony between stimuli (Kandil, Diederich, & Colonius, 2014). It has been proposed that stimulus onset asynchrony (SOA) of a stimulus pair has to fall within a certain time window for them to be integrated (Colonius & Diederich, 2004). The time window of integration (TWIN) model has been shown to efficiently describe temporal dependencies of crossmodal interactions (Colonius & Diederich, 2004; Kandil et al., 2014). Integration (and VE) rapidly declines for stimulus onset asynchronies outside of the window (Slutsky & Recanzone, 2001; van Wassenhove et al., 2007). The size of the TWIN depends on the experimental task and properties of the stimuli (Stevenson & Wallace, 2013). The temporal binding window is larger for speech than non-speech (e.g. flash and beep) stimuli - for perceptual fusion about 400 and

300ms respectively (Stevenson & Wallace, 2013). For the latter (non-speech stimuli) the TWIN is also asymmetrical - integration will still take place for larger asynchronies for visual stimulus preceding the auditory stimulus (Lewald & Guski, 2003; Stevenson & Wallace, 2013).

While spatio-temporal properties are critical for integration, they are not the only stimulus features that can influence multisensory processing. Semantic and synesthetic crossmodal correspondences have been shown to influence multisensory integration. The influence of crossmodal correspondences on multisensory integration was studied mainly by investigating multisensory enhancement (i.e. increase in accuracy or shorter reaction times). Stronger multisensory enhancement, as indexed by shorter reaction times (Diaconescu, Alain, & McIntosh, 2011; Laurienti, Kraft, Maldjian, Burdette, & Wallace, 2004; Molholm, Ritter, Javitt, & Foxe, 2004; Schneider, Engel, & Debener, 2008) and accuracy (Yuval-Greenberg, & Deouell, 2007), was observed for semantically congruent audio-visual stimuli. Similarly, more integration for synesthetically congruent comparing to incongruent stimuli was shown as indicated by smaller just noticeable difference in a temporal order judgement task (Parise & Spence, 2008). It has also been suggested that audio-visual semantic correspondences can facilitate learning (Barenholtz, Lewkowicz, Davidson, & Mavica, 2014; Meyerhoff & Huff, 2015).

Ventriloquist effect properties

The Ventriloquist illusion could be seen as a side effect of the multisensory gain coming from integration. It has been proposed that under the unity assumption, VE reduces perceptual conflict caused by the fact that one object cannot be present in more than one location (Bedford, 1999). As discussed above, visual signals provide more precise information about spatial location than auditory ones and for that reason it is optimal to rely more on the information from visual modality (Alais & Burr, 2004). This mislocalization of sound is directly related to audio-visual unity perception (Wallace et al., 2004), therefore the strength of the ventriloquist illusion in a given condition can be used to assess the level of multisensory integration.

A study by Choe and colleagues suggested that the ventriloquist illusion could be a result of a shift in response criterion, rather than in perception (Choe, Welch, Gilford, & Juola, 1975). Their claims however, have been criticised (Bertelson & Radeau, 1976, 1981) and a battery of research showing that VE is in fact a result of perceptual change has been published. First, VE occurs even when explicit instructions to ignore the visual stimulus are given and this persists despite subjects receiving feedback after each trial (Vroomen, Bertelson, & de Gelder, 1998). Second, the ventriloquist illusion was observed in experiments using staircase procedures, which would not allow for voluntary response adaptation (Bertelson & Aschersleben, 1998; Vroomen, Bertelson, & De Gelder, 2001). Participants were simply asked whether sound was presented in the left or right hemifield, and sound location on the next trial was adjusted based on the response (i.e. if sound on a previous trial was perceived on the right, sound on the following trial would be presented further towards the left) until response reversal. At the point of staircase reversal, observers are uncertain about stimulus location, which prohibits adopting post-perceptual strategies (Bertelson & Aschersleben, 1998). This sound localization procedure can be repeated for unimodal auditory and audio-visual presentations to measure ventriloquism. A difference in reversal points that is shifted towards the visual stimulus indicates occurrence of the perceptual illusion. Finally, spatially discrepant audio-visual presentations do not only cause VE, but also alter spatial perception in following unisensory presentations, causing a shift towards the previously presented visual stimulus (Bertelson, Frissen, Vroomen, & de Gelder, 2006; Frissen, Vroomen, de Gelder, & Bertelson, 2003; Lewald, 2002; Passamonti, Frissen, & Làdavas, 2009; Radeau, 1974). This phenomenon is called the ventriloquist aftereffect, which stems from crossmodal sensory recalibration and can be interpreted as adaptation to the changing acoustic environment (L. Chen & Vroomen, 2013).

By definition, the ventriloquist effect occurs for synchronous presentations (L. Chen & Vroomen, 2013), nevertheless crossmodal bias can also be observed for asynchronous presentations. The strength of the bias however, declines with increasing spatial and temporal discrepancy between visual and auditory stimuli (Hairston, Wallace, et al., 2003; Lewald & Guski, 2003; Slutsky & Recanzone, 2001). VE is smaller for peripheral than central locations (for the same spatial audio-visual disparity), which may be explained by the fact that for large eccentricities visual reliability decreases and auditory reliability increases (Charbonneau et al., 2013), therefore relying on visual information is no longer optimal.

Crossmodal bias is strongly correlated with reported common source perception (Hairston, Wallace, et al., 2003; Wallace et al., 2004). Interestingly, participants' sound localization seems to have a bimodal distribution, either showing a strong bias ~80-100% for trials where stimuli were perceived as a single event or no bias at all, or even repulsion effect (strongest for small spatial disparities) for trials where sound and visual were judged as separate stimuli (Wallace et al., 2004).

Audio-visual integration in ventriloquism has been suggested to be an automatic process that does not require deliberate attention. It has been shown that the strength of the ventriloquist effect is not influenced by either exogenous or endogenous attention (Bertelson, Vroomen, de Gelder, & Driver, 2000; Vroomen et al., 2001).

Despite the established position of crossmodal correspondences as a factor that modulates audio-visual integration, a few studies have suggested that they do not influence ventriloquist illusion. There was no difference in ventriloquism between pairs of voice and face or voice and flashing lights (Radeau & Bertelson, 1977), and no difference between VE for voice and upright/inverted faces (Bertelson, Vroomen, Wiegeraad, & De Gelder, 1994; Colin, Radeau, Deltenre, & Morais, 2001). At the same time, face inversion had a robust impact on the McGurk illusion (Bertelson et al., 1994; Colin et al., 2001). Modulatory effect of semantic congruency was shown however, in ventriloquist paradigm, in which talking faces were presented bilaterally (Kanaya & Yokosawa, 2011). A possible explanation of this could be that in the spatial ventriloquist paradigm with single image, the VE for incongruent images is already strong and additional binding cues (semantic congruency) cannot facilitate this effect any further. In Kanaya & Yokosawa's study, they used images of weaker saliency (faces presented bilaterally: one with moving lips shown and one with mouth area masked by a black oval), as a result, a positive influence of semantic congruency could be observed.

Neuroimaging studies investigating ventriloquism suggest that the illusion is related to perceptual change, which is reflected in brain activity. The mismatch negativity (MMN) is an ERP component that is classically elicited by presentation of a deviant stimulus e.g. sound played in a different location after a few consecutive presentations in the same location. It has been shown that a perceived auditory location shift can induce the MMN in a ventriloquist paradigm (Stekelenburg, Vroomen, & De Gelder, 2004). The evoked potential that has been identified as a signature of ventriloquism is the N260. In centroparietal electrodes the N260 amplitude is higher for illusion trials in the hemisphere contralateral to the visual stimulus. Significant differences in N260 amplitudes were observed for both hemispheres for illusion vs non-illusion trials and between hemispheres in illusion trials (Bonath et al., 2007). Source modelling showed that N260 originates from planum temporale, which was consistent with results of fMRI analysis conducted for the same paradigm that revealed analogous BOLD activity changes in planum temporale for illusion/non-illusion trials. The BOLD signal for illusion trials was reduced ipsilateral to visual stimulus, where sound was presented in central locations and an enhanced response was observed contralateral to the visual stimulus for laterally presented sounds. Greater relative contralateral/ipsilateral planum temporale activation reflected an illusory sound shift towards the contralateral visual stimulus (Bonath et al., 2014).

Awareness suppression methods

The next section of this chapter reviews the role of perceptual awareness in multisensory perception. Here I briefly describe most common techniques used to modulate perceptual awareness, to provide foundations for the coming section.

A very popular and relatively simple visual suppression technique is backward masking, in which a target stimulus is presented for a very short duration and immediately replaced with another stimulus - the mask (Kim & Blake, 2005). Backward masking has been shown to disrupt higher order visual processing (Noguchi & Kakigi, 2005). Other types of visual masking include forward masking, where the mask is presented before the brief target presentation (Huckauf & Heller, 2004), and forward-backward masking (also called sandwich masking), where the mask is presented both before and after the target (Harris, Wu, & Woldorff, 2011). In binocular rivalry, each eye is presented with a different image and observer perception alternates between the two, with only one image visible at a time (Logothetis, Leopold, & Sheinberg, 1996). Continuous flash suppression is a technique based on binocular competition. In CFS, one eye is presented with salient, high contrast masks (Mondrian rectangles) flickering with high frequency (10Hz), and the other eye with a lower contrast static image (Tsuchiya & Koch, 2005). Using CFS, a target picture can be suppressed from awareness for a prolonged duration comparing to binocular rivalry (Tsuchiya, Koch, Gilroy, & Blake, 2006). A variation of this technique is dynamic-CFS (dCFS), where original static Mondrians are replaced by dynamic gratings (Maruya, Watanabe, & Watanabe, 2008). The attentional blink is a phenomenon in which a second target presented shortly after the first in a rapid visual stream is not consciously perceived (Raymond, Shapiro, & Arnell, 1992). In visual crowding, a stimulus presented in periphery and surrounded by similar flankers can become unidentifiable (Huckauf & Heller, 2004). The above techniques are characterized by different strengths of suppression (for a review see: Breitmeyer, 2015; Kim & Blake, 2005). Breitmeyer proposed a functional hierarchy, where visual crowding and attentional blink are classed as weaker than masking, CFS and binocular rivalry (Breitmeyer, 2015). Interestingly though, he classed visual masking methods as stronger than CFS. Importantly, his analysis was based on comparison of results of separate studies using both different methods as well as different tasks/stimuli. A few studies in which those techniques were compared directly (for the same task and stimuli) reported either stronger suppression under CFS (Izatt, Dubois, Faivre, & Koch, 2014), or no difference between CFS and visual masking (Faivre, Berthet, & Kouider, 2012; Knotts, Lau, & Peters, 2018).

Finally, unconscious processes can be studied in patients with attentional or awareness disorders like blindsight or hemineglect. In blindsight, caused by lesions in primary visual cortex, patients preserve some discrimination ability, yet report unawareness of visual stimuli (Weiskrantz, 1986). Hemineglect is characterised by attentional deficits and lack of awareness of stimulation of one side of their body. Deficits are contralateral to lesions in parieto-temporal cortex (Kerkhoff, 2001; Vallar, 1998). Patients studies however, have major drawbacks as it is often not possible to compare conscious with unconscious perception and results cannot be generalized to the healthy population.

The above review focuses on visual masking methods, but similar techniques can be applied in other sensory modalities. For example, perceptual rivalry leading to bistable perception was shown for tactile stimuli (Conrad, Vitello, & Noppeney, 2012) and both backward and forward masking was successfully applied to auditory signals (Alves-Pinto, Baudoux, Palmer, & Sumner, 2010; Faivre, Mudrik, Schwartz, & Koch, 2014).

Multisensory perception and awareness

There seems to be a consensus among consciousness theories that awareness is closely related to the integration of environmental signals. It has been proposed that only integrated information can be available for conscious access (Crick & Koch, 1990; Grossberg, 1999; Newman & Baars, 1993; Thagard & Stewart, 2014; Tononi & Edelman, 1998). This stems from the assumption that awareness results from integrative processes in the brain and only coherent, unified percepts (of inputs that were already bound) can lead to the emergence of conscious states. On the other hand, Global Workspace Theory suggests that consciousness is necessary for the integration to occur (Baars, 1997, 2002, 2005; Dehaene & Naccache, 2001).

At the same time, signatures of multisensory integration have been observed at early stages of neural processing. Neuroimaging studies have shown that crossmodal binding takes place not only in higher order association cortices, but also in primary sensory cortices already (Foxe et al., 2002; Ghazanfar, Maier, Hoffman, & Logothetis, 2005; Lakatos, Chen, O'Connell, Mills, & Schroeder, 2007; Schroeder & Foxe, 2002). Experiments in non-human primates have shown direct connections between primary auditory and primary visual cortices (Cappe & Barone, 2005; Falchier, Clavagnier, Barone, & Kennedy, 2002), as well as inputs from visual cortex in somatosensory cortices and inputs in primary auditory cortex from somatosensory cortex (Cappe & Barone, 2005). Moreover short latency (<100ms post stimulus) electrophysiological evidence of multisensory interactions was observed in humans and non-human primates (Cappe, Thut, Romei, & Murray, 2010; Fort, Delpuech, Pernier, & Giard, 2002; Giard & Peronnet, 1999; Lin, Liu, Liu, & Gao, 2015; Senkowski, Talsma, Grigutsch, Herrmann, & Woldorff, 2007; Senkowski, Talsma, Herrmann, & Woldorff, 2005; Wang, Celebrini, Trotter, & Barone, 2008). These early stage multimodal interactions could mediate multisensory perception in the absence of awareness.

Multiple experiments suggested that a supraliminal stimulus from one modality can facilitate processing of a subliminal stimulus in another modality. Lip movements mismatching auditory speech took longer to break the CFS (Alsius & Munhall, 2013). An image of a hand matching actual hand position (proprioceptive signals) was reported to break the visual suppression in CFS paradigm faster than a mismatching image (Salomon, Lim, Herbelin, Hesselmann, & Blanke, 2013). Congruent vestibular information facilitated breaking the CFS of optic flow arrays (Salomon, Kaliuzhna, Herbelin, & Blanke, 2015). Moreover, visual awareness of a video was facilitated for matching soundtrack (e.g. moving train), but only when their presentations were synchronous, which suggests that this was not a result of semantic attentional priming. There was no difference in time to break the suppression between incongruent pairs and videos alone or for congruent pairs when the soundtrack was presented prior to the video (Cox & Hong, 2015). Similar interactions were shown for visual and haptic signals. Observers remained unaware of a moving dot for a longer time, when concurrent haptic stimulation (robotic arm touching their back) was in the opposite direction (Salomon et al., 2016). Congruent haptic gratings reduces suppression times of visual gratings comparing to incongruent and unimodal conditions (Lunghi, Lo Verde, & Alais, 2017).

Breaking CFS however, was questioned as a technique that measures unconscious processing, because differences in reaction times could potentially arise from processes happening after the stimulus has already reached awareness (Stein, Hebart, & Sterzer, 2011; Stein & Sterzer, 2014).

A couple of studies showed an increase in visual discrimination accuracy and the proportion of visibility reports depending on spatial congruency of audio-visual stimuli. Spatially aligned auditory stimuli improved visibility of a flashing dot under CFS (Aller, Giani, Conrad, Watanabe, & Noppeney, 2015). In a visual search task, presentations of collocated auditory cues increased elevation discrimination accuracy for masked visual targets (Ngo & Spence, 2010, 2012). Nevertheless, this could be a result of spatial attentional cueing toward exogenous auditory cues, and not of multisensory interactions.

Interestingly, researchers have also showed that perceptual awareness of subliminal stimuli can be affected by higher level crossmodal correspondences. Semantically congruent smells (of a rose or marker pen) prolonged dominance duration of corresponding images in binocular rivalry (Zhou, Jiang, He, & Chen, 2010). Similarly, sounds (bird or car) modulated visual dominance of a drawing depending on semantic congruency (Y.-C. Chen, Yeh, & Spence, 2011). In the attentional blink paradigm, if sound is presented synchronously with the second visual target it thrusts the image into awareness allowing for its identification (Olivers & Van der Burg, 2008). This phenomenon can be influenced by phonological correspondences between auditory and visual signals (Adam & Noppeney, 2014). Haptic presentations of a plexiglass grating had a similar effect on binocularly presented gabor patches with matching orientation (Lunghi & Alais, 2015; Lunghi, Binda, & Morrone, 2010). Bistable perception of touch was modulated by matching the direction of tactile vibration and apparent motion of a dot (Conrad et al., 2012). Direction of auditory motion also increased visual dominance of random dot kinematograms when the apparent motion of both stimuli was in the same direction (Conrad, Bartels, Kleiner, & Noppeney, 2010). In people that could read music, auditory melodies prolonged visual dominance of images of matching musical notes (M. Lee, Blake, Kim, & Kim, 2015).

The effects shown however, might not be specific to crossmodal interactions, as similar patterns were observed for experiments using only single modality. For example, it has been shown that an invisible stimulus can affect perception of a following visible stimulus and invisible priming in continuous flash suppression paradigm can enhance speed of visual categorisation (animal/not animal) (Koivisto & Rientamo, 2016). Familiarity of the image content itself can accelerate breaking Continuous Flash Suppression, which has been shown for recognizable words (Jiang, Costello, & He, 2007) and for upright comparing to inverted images (Jiang et al., 2007; Stein, Sterzer, & Peelen, 2012). Similar facilitation in breaking CSF can be observed for visual stimuli matching the one in visual working memory (Gayet, Paffen,

& Van der Stigchel, 2013), which suggests that merely semantic priming could be enabling faster access to perceptual awareness. Moreover, perception in binocular rivalry can be modulated by attention, which can extend the time that one of the pictures dominates (Chong, Tadin, & Blake, 2005), and the results of a recent study suggested that auditory enhancement of visual processing in binocular rivalry is merely due to salient attentional cues and not multisensory integration. It showed that silent gaps in sound presentation had the same influence on visual dominance in binocular rivalry as synchronous sounds (Pápai & Soto-Faraco, 2017).

Finally, all the experiments reviewed so far simply demonstrate that conscious signals from one sensory modality can influence perceptual awareness in the other modality. Thus, they neither provide evidence for multisensory integration for unaware signals (Deroy, Chen, & Spence, 2014) nor contradict Global Workspace Theory, as supraliminal signals from one modality should be globally accessible i.e. able to travel to other sensory areas and influence processing of subliminal stimuli in another modality.

There have been a few attempts to address both of those issues. Critically, GWT assumptions are challenged by experiments of Faivre and colleagues, which used subliminal audio-visual pairs as a prime in a congruency classification task. Semantic congruency of the prime modulated reaction times in classification of the following supraliminal audio-visual target, but only if audio-visual relationships were previously learned (Faivre et al., 2014). This congruency effect however, does not imply multisensory integration and could result from comparison of auditory and visual signals. A recent study has shown multisensory enhancement for unconscious visual signals in a patient with Posterior Cortical Atrophy, who was unable to perceive visual stimuli of short duration (Barutchu, Spence, & Humphreys, 2018). Importantly however, this multisensory gain was only shown for reaction times, but not for detection accuracy. In addition, no difference was observed between semantically congruent and incongruent audio-visual pairs. Finally, the reduction in reaction times was only significant for auditory stimuli either preceding or presented synchronously with visual but disappeared for stimulus onset asynchrony greater than 50ms when the visual stimulus was leading (i.e. "multisensory loss"). This implies that conscious auditory perception was necessary for the gain to occur. Taken together, this suggests that the multisensory enhancement in this paradigm could nevertheless result from attentional crossmodal priming. One previous study suggested that VE can occur in patients with spatial hemineglect, showing shift of perception towards undetected targets (Bertelson, Pavani, Ladavas, Vroomen, & de Gelder, 2000). This shift, however, was only present in patients' neglected, but not intact hemifield, therefore these results should be interpreted with caution. Moreover, as it was only shown in patients, the effect could be specific to the disorder and not unconscious processing in general.

Interestingly, observers were able to use congruency of audio-visual speech cues, where lip movements were not consciously perceived (under continuous flash suppression), to learn their predictive statistical pattern and utilize it in a concurrent target recognition task (Palmer & Ramsey, 2012). At the same time, the McGurk effect was abolished when visual lip movements were presented under CFS (Ching, Kim, & Davis, 2019; Palmer & Ramsey, 2012). This suggests that crossmodal integration might not be possible for unconscious signals, even though certain features of those signals can be extracted.

Given the above concerns and contradictory findings, further studies are needed to investigate the interplay between awareness and multisensory perception. First, to examine the extent to which sensory areas can communicate without awareness and what types of subliminal information can travel through the global network. Both integrative and non-integrative multimodal paradigms could be used to test Global Workspace Theory predictions. Nevertheless, unconscious processing of multimodal signals can be different when they are integrated. Therefore, further research is necessary to provide unambiguous evidence of crossmodal integration for unaware stimuli, if they are indeed possible. Those issues are examined in Chapters 3 to 5 of the current thesis.

Multisensory metacognition

Metacognition is sometimes described as "cognition about cognition" and refers to observer's knowledge about the quality of their perception (i.e. sensory uncertainty) or accuracy of their perceptual decisions. For example, to asses metacognition in a picture discrimination task, observers would not only be asked to report which picture was presented (this is a so-called first order task), but additionally to state how confident they are about their choice (second order task).

Metacognition, similarly to awareness (Deroy et al., 2014; Faivre, Arzi, Lunghi, & Salomon, 2017), was mainly studied in unisensory paradigms, with the vast majority of experiments conducted in the visual domain (Faivre et al., 2017). With respect to various sensory modalities, it has been investigated whether metacognitive performance is correlated for different senses i.e. supramodality of metacognition. So far, such a relationship has been shown for vision and audition (Ais, Zylberberg, Barttfeld, & Sigman, 2016; De Gardelle, Le

Corre, & Mamassian, 2016), vision and touch, audition and touch (Faivre, Filevich, Solovey, Kühn, & Blanke, 2018) and vision and warmth perception (Beck, Peña-Vivas, Fleming, & Haggard, 2019). Until now, only one study examined metacognition in a multisensory task, showing similar correlations between bimodal (audio-visual) and unimodal (auditory and visual) metacognitive efficiency (Faivre et al., 2018). Since they specifically focused on a non-integrative case of multisensory perception, metacognition in the context of multisensory integration remains unexplored.

In a recent opinion paper, Deroy et al. put forward two interesting questions regarding causal metacognition, i.e. confidence about the causal structure of multimodal events that could originate from one common or two separate sources (integration vs segregation) (Deroy, Spence, & Noppeney, 2016). First, they considered whether observers have metacognitive access to uncertainty about their final multisensory percept, individual unisensory uncertainties of the stimuli, or both. Second, they considered what level of metacognitive access is present for multisensory illusions (e.g. McGurk effect, Ventriloquism) that can create perceptual metamers. If subjects would be more confident about non-illusion than illusion trials leading to the same perceptual choice, that would suggest access to additional information about the process of causal inference. Both problems are addressed by the study described in Chapter 6.

CHAPTER 2. METHODOLOGICAL FOUNDATIONS

This chapter describes the methodological foundations of subsequent experimental chapters. It includes adaptive methods applied to identify visibility thresholds in CFS studies (Chapters 3 and 4), techniques used to quantify the ventriloquist effect across the thesis, and observers' metacognitive abilities in Chapter 6 and the basis of multivariate pattern analysis applied to EEG data in Chapters 4 and 6.

Measuring ventriloquism

There are multiple approaches to the evaluation of the ventriloquist illusion. Crucially, analysis of VE depends on the task that participants are asked to perform, which can either focus on sound localization, spatial alignment of the stimuli or unity perception. Sound location reports allow for direct assessment of visual influence on auditory perceptions and for that reason this task was used in experiments in Chapters 3-5. Moreover, some of the VE assessment methods are only suitable for quantitative and not qualitative response analysis.

One of the simplest ways to quantify ventriloquism is computing spatial displacement between real and perceived sound locations expressed in visual degrees (Bertelson & Radeau, 1981). Such displacement defines the absolute bias of vision on audition (Bertelson & Radeau, 1981). As the size of this shift is not informative about strength of the illusion, without knowing audio-visual disparity a measure that accounts for it was introduced by Welch & Warren (Welch & Warren, 1980). So called relative crossmodal bias (CMB) is computed as:

$$(1)\ CMB = \frac{A_{resp} - A_{true}}{V_{true} - A_{true}} \times 100\,,$$

where A_{resp} is reported sound location, A_{true} real sound location and V_{true} is visual signal location. Computation of relative CMB is simple and easy to interpret, as a CMB greater than zero indicates occurrence of the ventriloquist illusion. However, this interpretation becomes problematic in paradigms with a discrete, finite number of responses. For example, if subjects can only report sound as coming from far left, left, right or far right, CMB for the extreme locations (far left/right) would always be greater than zero as a result of noise in auditory estimates. Sound presented in the far left location can be either reported correctly or biased towards the right. Unavailability of a "further left" location creates an artificial shift in perception, where perception beyond the "far left" location simply cannot be reported. Similarly, CMB is prone to individual perceptual biases in sound localization (i.e. sound presented in the middle can be more often perceived as coming from the right than as coming from the left or the other way around). Because of this, CMB is not an appropriate measure to prove that observers have experienced the ventriloquist illusion.

The aim of Chapter 3 was to test whether VE can occur for invisible flashes, therefore it was necessary to find a bias free measure. First, we turned to the linear regression model, which predicts sound location perception depending on auditory and visual signals:

$$\text{(2)}\quad \mathrm{A}_{\mathrm{resp}} = \mathrm{A}_{\mathrm{true}} * \beta_{\mathrm{A}} + \mathrm{V}_{\mathrm{true}} * \beta_{\mathrm{V}} + \mathrm{C}$$

where again A_{resp} is reported sound location, A_{true} and V_{true} are true auditory and visual locations, and β_{A} and β_{V} are the estimated auditory and visual regression coefficients respectively (Aller et al., 2015). The main advantage of this approach is that it is bias free as under the null hypothesis (no visual influence) the expectation of the visual parameter estimate should be equal to zero (any general bias such as pointing more often towards the right would be contained in the constant C).

Applying the above model to experiments in Chapters 3-4 has proven to be more complex than originally anticipated. Crucially, the assumption was that visual influence on audition would differ for visible and invisible stimuli (i.e. separate $\beta_{V\_visible}$ and $\beta_{V\_invisible}$ coefficients), however we did not hold assumptions regarding auditory coefficients or response bias and whether they should be each fitted together or separately for different visibility conditions. Moreover, we wanted to asses neural ventriloquism (i.e. based on classifier predictions; details in Chapter 4) for illusion and non-illusion trials, which again introduced two additional factors into the analysis. Without having clear assumptions regarding auditory coefficients in such models, this introduced the problem of model selection, as between 1 and 4 auditory coefficients could be fitted in 4 separate models. Given that model comparison criteria (BIC, AIC and Log Likelihoods) did not provide clear answers as to which model best fits the data, and that the creation of optimal computational model of auditory processing in ventriloquism was not one of the aims of this thesis, we decided to look for a method that would allow us to avoid the unnecessary complexity.

Finally, to quantify the ventriloquist effect, we simply compared sound perception when the visual stimulus is presented on the right and on the left. For each true auditory location, we computed the difference between mean reported sound location for visual right and left conditions. This difference, averaged over all sound locations (N), served as an index of ventriloquism:

$$(3)\ VE = \frac{\sum \bar{A_i}_{VR} - \bar{A_i}_{VL}}{N},$$

where $\bar{A}_{i\,VR}$ is mean reported auditory location for true auditory location i and visual location right (VR) or left (VL). For illustration see Figure A.2, Appendix A. This approach is not only bias free, but also easily applicable to paradigms used in Chapters 3 and 4.

To evaluate the ventriloquist effect for qualitative reports - e.g. common source judgement task - one can compare the proportion of common source/unity reports for different conditions, i.e. dependent on spatial or temporal congruency (Slutsky & Recanzone, 2001). Similar analysis can be performed for sound localization task responses, e.g. sound localization accuracy (Bertelson, Pavani, et al., 2000), but this is not informative for continuous responses (or discrete responses with multiple levels) as incorrect judgement could also mean a shift in the direction opposite to where visual stimulus was presented. Such analysis works best for tasks with two alternative choices, e.g. left vs right, common vs separate sources etc. Modulation of spatial localization accuracy (or fraction of unity reports) by spatial alignment of audio-visual stimuli indicates multisensory integration. This approach was used in experiments in Chapter 5, where subjects only discriminated between left and right sound locations, and in Chapter 6, where participants reported sound and picture as coming from the same or different locations.

Adaptive staircase method

Adaptive staircase is a psychophysical technique that can be used to achieve the desired level of task performance by titrating stimulus parameters. Adaptive methods are frequently used in consciousness studies to find thresholds of subliminal presentation (e.g. Rothkirch & Hesselmann, 2018), in metacognition to match performance between tasks (e.g. Fleming, Huijgen, & Dolan, 2012) and in cognitive science in general to optimize efficiency of experimental procedures (Kingdom & Prins, 2010). A variety of adaptive procedures can be used to achieve this aim (for an overview see: Kingdom & Prins, 2010; Leek, 2001), but the review below is limited to the methods used in the experimental chapters of this thesis.

Adaptive procedures were first introduced into auditory research (Békésy, 1947; Hughson & Westlake, 1944) and developed by Dixon and Mood (Dixon & Mood, 1948). The concept is rather straightforward: to find a threshold at which an event would occur (e.g. yes answer), stimulus intensity is adjusted after each trial, i.e. it is decreased when the event occurs or is increased when it does not (no answer) (Cornsweet, 1962). This basic version of a staircase procedure is called an up-down method and it equates to 50% probability of the event. To find thresholds for different probabilities of the event/task performance, the original procedure has to be modified. In a transformed up-down method, this is achieved by changing the number of consecutive yes responses after which stimulus intensity decreases. A staircase in which intensity decreases only after two consecutive yes answers and increases after a single no answer would be called one-up/two-down procedure and equates to about 70.7% accuracy (Levitt, 1971). Another way to modify the traditional staircase method to achieve desired performance is to change relative values by which the intensity is decreased and increased. This method is called weighted up-down method, and relative step sizes can be computed as:

$$(4) \quad \frac{\Delta_-}{\Delta_+} = \frac{100 - Acc}{Acc} \ ,$$

where $\frac{\Delta_-}{\Delta_+}$ is step down to step up ratio and Acc stands for target accuracy in percent (Kaernbach, 1991).

In the simple one-up/one-down staircase, changing stimulus intensities can become apparent to the observer – especially for intensities close to the threshold, but this issue can be minimized using interleaved staircases (Ehrenstein & Ehrenstein, 1999). In the interleaved procedure, two staircases can be alternated every trial or randomly.

Before using adaptive methods, in addition to required performance and procedure type (i.e. one-up/two-down), one must determine i. starting value, ii. step size, iii. when to terminate the procedure (stopping/convergence criterion) and iv. how to compute the threshold (Cornsweet, 1962). Ideally, initial stimulus intensity should be close to the threshold value, so the staircase would converge relatively fast. For interleaved staircases, usually values above and below the expected threshold are chosen as starting points. Choice of step size has to be suited to the purpose: for larger step sizes staircase will converge faster, but estimated threshold will be less precise. Staircase termination is usually set after a certain number of trials or reversals (changes from one answer to another). The appropriate way to compute the threshold largely depends on the convergence criterion/ used step size. The threshold can be calculated either as mean for set numbers of trials (trials at the start of the staircase should be excluded) or mean for set number of reversals (again, first few reversals should be discarded) e.g. last 10 trials or last 5 reversals (Kingdom & Prins, 2010).

Signal Detection Theory and metacognition

Signal Detection Theory describes perceptual decision making and can be applied to any discrimination task with two options (this can include: yes/no paradigms, detection: signal present/absent, and N-alternative forced choice tasks) (Kingdom & Prins, 2010). SDT assumes that evidence about the signal (decision variable) follows a normal distribution and variance of both signals are equal. An observer makes a perceptual choice based on whether evidence available on a given presentation exceeds a certain threshold - a decision criterion (solid line in Figure 2.1).

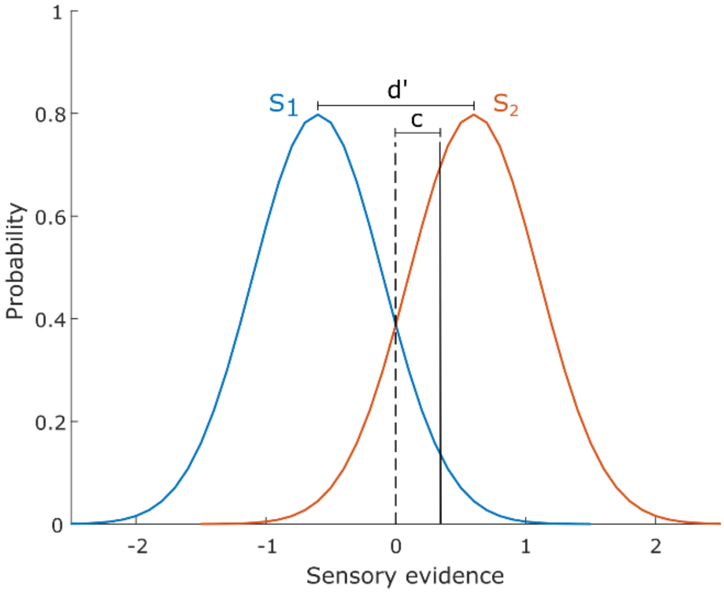

Figure 2.1 Probability distribution of sensory evidence for two signals in SDT model. D' is a difference between means of the distribution. The solid line marks the decision criterion and the dashed line the optimal decision criterion - distance between them is the response bias c. Figure based on: (Hautus, 2015; Stanislaw & Todorov, 1999).

Task performance in the SDT model is described by d-prime, which is equal to the difference between means of evidence distribution for noise and signal in a detection task or between two signals in a discrimination task. D-prime can be computed as the difference between z-transformed (inverse-normal transform) hit rate and false alarm rate (for SDT overview see: (Hautus, 2015; Stanislaw & Todorov, 1999)):

$$(5)\ d' = z(HR) - z(FR)\,,$$

where HR- hit rate, FR – false alarm rate. Hit rate is the proportion of yes responses in signal present trials and false alarm rate is the proportion of yes responses in signal absent trials (see Table 2.1). Response bias is a tendency to report one of the signals more often than the other (reporting presence/absence of the signal more often in detection task). Response bias is the difference between the optimal decision criterion (marked as dashed line on Figure 2.1) and the actual decision criterion (solid line). The receiver operator characteristic (ROC) curve describes the relationship between proportion of hits and false alarms for given d’, depending on decision criterion c adopted by the observer (Kingdom & Prins, 2010).

		Response	
		S1	**S2**
Trial Type	**S1**	Hit (True positive)	Miss (False negative)
	S2	False Alarm (False positive)	Correct Rejection (True negative)

Table 2.1 First order task judgements.

Metacognition is the ability to monitor one’s own performance. Discrimination between two signals is a first order task; if the observer would be also asked to rate their confidence this would be a second order task. Measuring performance in second order tasks is much more challenging. Early metacognition studies used correlation measures to assess relationship between confidence and first order task performance (Nelson, 1984). Later research however, argued in favor of SDT approaches to avoid confounds related to response bias (Barrett, Dienes, & Seth, 2013; Galvin, Podd, Drga, & Whitmore, 2003; Maniscalco & Lau, 2012, 2014). First order task sensitivity, measured by d’, describes the ability to discriminate between the two signals. Type 2 sensitivity describes the ability to discriminate between own correct and incorrect judgment. Technically, for high vs low confidence judgement, it is possible to compute type 2 d’ using type 2 hit and false alarm rates (see Table 2.2). Unfortunately, evidence distributions for confidence judgements are not gaussian, which violates SDT assumptions (Galvin et al., 2003). It has been proposed that metacognitive performance could be assessed using type 2 ROC curves (for review see: Barrett et al., 2013). However, type 2 ROC curves are dependent on the type 1 model parameters (Galvin et al., 2003). Maniscalco and Lau have recently introduced a model, which allows computing meta d', which is defined as a d' that a metacognitively optimal observer would have in the type 1 task, assuming the empirically derived response criterion, which would produce collected type 2 responses (Maniscalco & Lau, 2012). This is achieved by fitting the maximum likelihood estimation model, which finds the meta d' parameter maximizing the likelihood of obtaining type 2 data, given the type 1 decision criterion. This meta d' is still dependent on type 1 response criterion, but meta d'/d' ratio (metacognitive efficiency) describes purely metacognitive ability. A metacognitive efficiency of 1 would mean that the observer is metacognitively optimal and utilizes all available information for the confidence judgement.

		Confidence	
		High	**Low**
Type 1 response	**Correct**	Type 2 Hit	Type 2 Miss
	Incorrect	Type 2 False Alarm	Type 2 Correct Rejection

Table 2.2 Second order task judgements.

Multivariate pattern analysis of EEG data

Over the past two decades machine learning techniques are increasing in importance in neuroimaging studies. There are two categories of machine learning: unsupervised and supervised. Unsupervised learning does not require any description of the data or its features, but can be used for grouping subjects, examples or chosen feature sets based on their similarity. Supervised methods are generally more commonly used in cognitive neuroscience in order to identify brain activity related to specific conditions, but unsupervised methods are also successfully used, e.g. to identify functional regions in fMRI data (Filzmoser, Baumgartner, & Moser, 1999).

Supervised machine learning was widely applied in Brain Computer Interfaces long before it was applied in cognitive neuroscience research. Pioneering BCI papers were published in the early nineties and focused on its application as a communication tool for disabled patients (Pfurtscheller, Flotzinger, & Kalcher, 1993; Wolpaw, McFarland, Neat, & Forneris, 1991). The concept of utilizing brain activity for communication is even older as it had already been put forward in the seventies (Vidal, 1973). The first BCIs employed EEG, because compared to other neuroimaging methods (fMRI, PET,MEG) EEG equipment is cheap and portable, which makes it most suitable for the purpose – communication with patients (e.g. with locked in syndrome) on a daily basis (Wolpaw et al., 2000). Interestingly, it is fMRI that paved the way for multivariate approaches in cognitive neuroscience (for historical overview see: O'Toole et al., 2007).

In supervised machine learning, an algorithm is trained to predict labels (e.g. experimental conditions) of the data. The choice of algorithm depends on the data type. A regression model is needed to predict labels of continuous character, e.g. luminance of visual stimulus and classification algorithm to discriminate between two or more conditions (classes) e.g. birds and dogs. Multivariate pattern analysis of neuroimaging data is advantageous compared with traditional univariate approach as it is characterised by greater sensitivity (review in: Pereira, Mitchell, & Botvinick, 2009), therefore allowing to find neural representations of specific cognitive states. For example, in traditional event related potential (ERP) analysis, ERP components are compared between conditions for individual channels. In addition, MVPA can utilize information of specific patterns of activity and dependencies between individual EEG channels. If a classifier can successfully discriminate between mental states, it is then possible to identify where (in fMRI) or when (in EEG, MEG) their neural representations differ.

Below I outline the process of decoding analysis, mainly focusing on application to EEG data.

The first step of the procedure is EEG preprocessing. Raw EEG data can also be used for MVPA analysis, but there are a few issues to consider. Importantly, artifacts and eye movements should be removed not only to reduce noise, but also to avoid confounds. Assume one wants to discriminate between left and right visual stimuli. If an observer would look towards the stimuli, the classifier could use eye movement information instead of neural spatial representations, which are of interest, for label prediction (n.b. in such experiments, observers are asked to fixate on the centre of the screen, but even occasional movements could be picked up by the classifier). In order to increase the signal to noise ratio, data can be downsampled, or sliding time windows can be used for classification (Grootswagers, Wardle, & Carlson, 2017).

The next step involves the creation of examples and feature selection for analysis (see Figure 2.2). In EEG, each trial can be considered independent, so single-trial decoding is a good option with the most straightforward interpretation. Examples can also be created by trial averaging, which can increase the signal to noise ratio, but it reduces the number of examples and can be a source of confounds, e.g. for averaging trials of identical stimuli, but with different behavioural reports or perception. Another way to increase classifier performance is to select the most informative features for the analysis, which can be done using unsupervised, data-driven techniques such as Principal Component Analysis or by restricting decoding analysis to a region on interest (ROI) based on previous studies (overview in: Grootswagers et al., 2017). Crucially, the choice of ROI should not depend on classifier performance within that region (i.e. it should be selected either a priori or in cross-validated fashion: optimized on part of the data and applied to the rest) (Skocik, Collins, Callahan-Flintoft, Bowman, & Wyble, 2016). Dimensionality reduction is most beneficial in neuroimaging techniques, like MEG or fMRI, that have a large number of sensors/voxels, increasing the number of possible features. In EEG analysis, a feature can be an amplitude value at a given channel and time point (downsampling can also be considered a dimensionality reduction). Another possibility would be using power-frequency data as features.

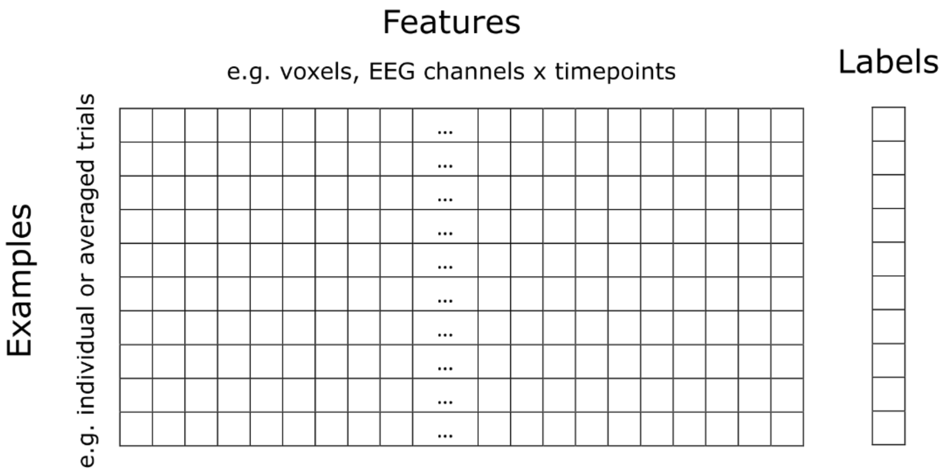
Features
e.g. voxels, EEG channels x timepoints
Labels
Examples
e.g. individual or averaged trials
...

Figure 2.2 Dataset for MVPA analysis. Each row represents an example e.g. single trial, each column represents a single feature e.g. EEG channel for a single timepoint. Each example should have an assigned label: either a continuous value (for regression analysis) or a category (for classification).

Before decoding analysis, an appropriate algorithm must be chosen. Linear algorithms are most commonly used in neuroimaging studies as non-linear ones can lead easily to overfitting in small sample sizes and decrease generalizability of the findings (Lemm, Blankertz, Dickhaus, & Müller, 2011). The choice between regression and classification depends mostly on the type of data – for continuous labels regression is the only option and classification is the clear choice for categorical stimuli. For a few discrete levels of stimulation, e.g. spatial locations, one can consider using regression or multiclass classification, where the decision would depend on the experimental question, i.e. whether perception of the stimuli between the used levels is likely to occur (this is definitely true for auditory perception, less so for visual unless the signal is noisy). Both approaches are used in such cases (e.g. (Aller & Noppeney, 2019) for regression, (Salti et al., 2015) for classification). Some of the most popular algorithms include k-nearest neighbours (knn), Gaussian Naive Bayes (GNB), linear discriminant analysis (LDA) and support vector machines (SVM). The nearest neighbour classifier is one of the simplest, not computationally heavy classifiers, but not extremely powerful. It is a discriminative algorithm that finds the label by choosing the most common class among the k most similar examples in the training dataset (i.e. nearest neighbour). Its discriminative performance can be improved when it is used in conjunction with some dimensionality reduction method (Pereira et al., 2009). The Naïve Bayes algorithm uses the conditional probability of features given the label in the training data, to estimate the probability of a label given features in the testing data (Bruce & Bruce, 2017). LDA is a generative model, that tries to find a linear combination of features, which maximizes variance between classes and minimizes the variance within classes (Bruce & Bruce, 2017). SVM is a discriminative model, that creates a hyperplane – decision boundary, which best separates the features between two classes (Hastie, Tibshirani, & Friedman, 2009). Some classifiers (e.g. LDA) have the inherent ability to discriminate between multiple classes, some would perform one vs all classification to find the best fit for the data in multiclass classification.

The next step is training the classifier on part of the data and using it to predict labels on another part of the data. Training and testing data sets must be independent to allow for evaluation of classifier performance and its generalizability to the population. If the classifier would have access to all the data at the training stage (so-called double dipping) it would not give a reliable performance estimate as classifier performance would always be better than chance, even for features containing only noise. The process that enables assessment of classifier performance is cross-validation. Cross-validation is a procedure in which data is divided into training and testing sets. Data, at least in the training set, should be balanced, so it includes equal numbers of examples from each class to make reliable predictions. The testing set does not have to be balanced (and it might be useful to have predictions for all trials for subsequent analyses), however in such a case balanced accuracy should be computed to asses classifier performance (mean prediction accuracy for each class averaged over all classes). At this stage, one should also consider possible confounds and take those into account when balancing trial numbers (Snoek, Miletić, & Scholte, 2019). There are a few commonly used cross-validation (partitioning) schemes. In leave-one-example-out cross-validation, the classifier is trained on all, but one sample, and tested on that sample – the process is then repeated until all samples have predictions. This partitioning scheme allows us to maximize the number of examples used for training (which can improve performance, see e.g.: Grootswagers et al., 2017) but is computationally expensive. In n-fold cross-validation, data is divided into n equal folds, where n-1 folds are used for training and the remaining fold for testing. The procedure is then repeated for each fold. In split-half cross-validation, data is simply divided into two equal parts, where one is used for training and the other for testing. In fMRI experiments, leave-one-run-out cross-validation is popular, as trials within a single run are not independent (so trials from that run can be used either for training or for testing, but not both). In EEG studies, each trial can be considered independent, therefore any partitioning scheme can be easily applied. To reduce data variability, normalization can be applied to each channel/voxel. Importantly, this should be done based on normalization parameters from the training data sets that are then applied to both training and testing data (Schmack, Burk, Haynes, & Sterzer, 2016). Therefore, normalization should be applied at the cross-validation stage (that is why it is described in this section, even though it could be considered a preprocessing step). Common normalization methods include: subtraction of the mean, z-score normalization, setting data to have values ranging from zero to one (Lu, Zhang, Xu, & Liu, 2018; Pereda, Quiroga, & Bhattacharya, 2005; Schmack et al., 2016; Sterzer, Haynes, & Rees, 2008).

The final stage of the analysis is testing classifier accuracy against chance performance. In ROI analyses, simple parametric tests can be used (Bode & Haynes, 2009), but in recent years, researchers argued in favour of permutation based approaches (Pereira & Botvinick, 2011; Stelzer, Chen, & Turner, 2013). Within subject permutation tests can be performed using accuracies obtained from repeated classification of the same data with shuffled target labels. Across subject permutation tests can use either subject level performance or individual null distributions (created by decoding random labels) to create group level null distribution. Analyses over time, or for multiple voxels, additionally require multiple comparison correction (MCC). As consecutive timepoints and neighbouring electrodes in EEG data or neighbouring voxels in fMRI data are not independent, cluster based approaches (that make use of temporal or spatial proximity information) to MCC are commonly used (Aller & Noppeney, 2019; Grootswagers et al., 2017; Stelzer et al., 2013). One limitation of cluster analysis is having to set an arbitrary threshold for combining features into clusters (Oosterhof, Connolly, & Haxby, 2016). Threshold Free Cluster Enhancement (TFCE) method, introduced by Smith and Nichols, solves this problem by computing scores for each data point using statistical maps thresholded at different levels (Smith & Nichols, 2009). In TFCE, both signal intensity and supporting information from neighbouring data points (for spatial or temporal neighbours) are used to optimize sensitivity. This method has been successfully applied to both EEG and fMRI data (Mensen & Khatami, 2013; Salimi-Khorshidi, Smith, & Nichols, 2011).

CHAPTER 3. INVISIBLE FLASHES ALTER PERCEIVED SOUND LOCATION

Patrycja Delong, Máté Aller, Anette S. Giani, Tim Rohe, Verena Conrad, Masataka Watanabe, Uta Noppeney

Contributions:

PD and UN designed the study. PD acquired the data. PD analysed the data, under the supervision of UN. MA, ASG, VC, TR, MW contributed to the design, data acquisition or analysis of a prior unpublished study that motivated and guided this study. All authors wrote the manuscript and approved the final version of the paper for submission.

This chapter was published as:

Delong, P., Aller, M., Giani, A. S., Rohe, T., Conrad, V., Watanabe, M., & Noppeney, U. (2018). Invisible Flashes Alter Perceived Sound Location. Scientific Reports, 8(1), 12376.

Results of this chapter were presented at:

1. BNA 2017 Festival of Neuroscience, Birmingham, UK 2017

Poster: “The invisible ventriloquist – can unaware flashes alter sound perception?”

2. The 22nd meeting of the Association for the Scientific Study of Consciousness, Krakow, Poland, 2018

Talk: “The invisible ventriloquist: audio-visual integration in the absence of perceptual awareness”

Abstract

Information integration across the senses is fundamental for effective interactions with our environment. The extent to which signals from different senses can interact in the absence of awareness is controversial.

Combining the spatial ventriloquist illusion and dynamic continuous flash suppression (dCFS), we investigated in a series of two experiments whether visual signals that observers do not consciously perceive can influence spatial perception of sounds. Importantly, dCFS obliterated visual awareness only on a fraction of trials allowing us to compare spatial ventriloquism for physically identical flashes that were judged as visible or invisible.

Our results show a stronger ventriloquist effect for visible than invisible flashes. Critically, a robust ventriloquist effect emerged also for invisible flashes even when participants were at chance when locating the flash.

Collectively, our findings demonstrate that signals that we are not aware of in one sensory modality can alter spatial perception of signals in another sensory modality.

Introduction

Information integration across the senses is critical for effective interactions with our natural environment. The extent to which multisensory integration depends on perceptual awareness is controversial (Deroy et al., 2014; Deroy, Faivre, et al., 2016; Faivre et al., 2014; Mudrik, Faivre, & Koch, 2014). According to the global neuronal workspace (GNW) model, consciousness relies on information being broadcast via long-range connectivity in a frontoparietal system (Baars, 2005). As a result, signals that we are aware of in one sensory modality should be able to influence processing in brain areas dedicated to processing signals from another sensory modality. By contrast, processing of signals that we are not aware of should be largely confined to their own sensory system and have only little effect on perception of signals in another sensory modality.

Indeed, in line with the first prediction, a vast body of research has demonstrated that aware signals from one sensory modality thrust unaware signals in another sensory modality into perceptual awareness according to the classical multisensory principles of temporal coincidence, spatial concordance and semantic and phonological congruency (Adam & Noppeney, 2014; Alsius & Munhall, 2013; Y.-C. Chen et al., 2011; Conrad et al., 2010, 2013; Faivre et al., 2017; Lunghi & Alais, 2015; Lunghi et al., 2010, 2017; Salomon et al., 2015, 2013; Zhou et al., 2010). With respect to the spatial ventriloquist illusion, we have recently demonstrated that a sound that we are aware of can boost a flash under dynamic flash suppression into perceptual awareness depending on audiovisual spatial congruency (Aller et al., 2015).

By contrast, little evidence has been provided for modulatory effects of unaware signals in one sensory modality on aware signals from another sensory modality. Most notably, the McGurk illusion has been shown to be abolished when visual facial movements are obliterated from awareness under flash suppression (Palmer & Ramsey, 2012) or in bistable perception (Munhall, ten Hove, Brammer, & Paré, 2009).

Surprisingly, a recent study demonstrated that participants were faster at responding to supraliminal audiovisually congruent (resp. incongruent) stimuli when those supraliminal stimuli were preceded by subliminal congruent (resp. incongruent) primes (Faivre et al., 2014).

Yet, while these results suggest that the brain can compare auditory and visual letters/phonemes in the absence of awareness, congruency priming does not necessarily imply genuine multisensory interactions. Further, the effects were only observed in terms of response times rather than perceptual representations or choices.

To our knowledge, only one previous study provided tentative evidence that unaware visual signals in patients with hemi-neglect induce a ventriloquist effect and bias patients' perceived sound location (Bertelson, Pavani, et al., 2000). These results, however, need to be interpreted with caution, as the ventriloquist effect was reported as significant for visual signals only in patients' neglected, but not in their intact hemifield. Furthermore, this study characterized the ventriloquist effect only for unaware but not for aware visual signals in patients' neglected hemifield.

In light of these controversial findings it remains unknown whether unaware signals in one sensory modality can influence conscious perception of signals in another sensory modality. Given accumulating evidence that multisensory interactions emerge already at the primary cortical level (Bizley, Nodal, Bajo, Nelken, & King, 2007; Bonath et al., 2014; Falchier et al., 2002; Ghazanfar & Schroeder, 2006; H. Lee & Noppeney, 2014; Rohe & Noppeney, 2016) one may argue that potentially low-level spatiotemporal information rather than phonological information as in the McGurk illusion may be integrated in the absence of awareness.

Combining the spatial ventriloquist illusion (Bertelson & Radeau, 1981; Wallace et al., 2004) and continuous dynamic flash suppression (dCFS) (Maruya et al., 2008) we investigated in two psychophysics experiments whether visual signals that observers did not consciously perceive can influence spatial perception of sounds. Critically, we adjusted the saliency of the visual flash, such that the dynamic continuous flash suppression obliterated visual awareness only in a fraction of trials. This allowed us to compare spatial ventriloquism for physically identical flashes that do or do not enter participant's awareness.

Methods

Participants

After giving informed consent, 41 healthy young adults (34 females, 39 right-handed, mean age: 20.1 years, standard deviation: 4.1, range: 18-41) participated in experiment 1, 28 subjects (22 female, 27 right handed, mean age: 19.3 years, standard deviation: 1.4, range: 18-25) in experiment 2. The study was performed in accordance with the principles outlined in the Declaration of Helsinki and was approved by the local ethics review board of the University of Birmingham.

For the first experiment we hypothesized a medium effect size (Cohen's d=0.5) for the ventriloquist effect in the invisible condition. Hence, we computed sample size (n) for a one sided t-test and desired statistical power equal to 0.9, n=35. To determine sample size for experiment 2, we used an effect size based on the sample from the first study (Cohen's d≈0.7); for the same statistical power (0.9) we obtained n=18. We continued with data acquisition until the number of included data sets was equal to the required sample size (i.e. excluded subjects were replaced; see section exclusion criteria).

Stimuli and apparatus

Participants sat in a dimly lit room in front of a computer monitor at a viewing distance of 95cm. They viewed one half of the monitor with each eye using a custom-built mirror stereoscope. Visual stimuli were composed of targets and masks that were presented on a grey, uniform background with a mean luminance of 15.6 cd/m^2. On the 'flash present' trials, one eye viewed the target stimulus (i.e. the flash), which was a grey disc (Ø 0.3°) presented for 50ms in the upper left, lower left, upper right or lower right quadrant, i.e. at ± 3° visual angle along the azimuth and ± 1.2° elevation from a grey central fixation dot. The elevation of ± 1.2° was selected to enable effective multisensory interactions between flash and sound irrespective of flash elevation. The luminance of the flash was adjusted individually via adaptive staircases to obtain 60% invisible trials. To suppress the flash's perceptual visibility, four dynamic Mondrians (Ø 2.08°, mean luminance: 48 cd/m^2) were shown to the other eye (Maruya et al., 2008). In dynamic CSF, original static rectangles (Tsuchiya & Koch, 2005) are replaced with dynamically moving gratings (Aller et al., 2015; Maruya et al., 2008). The Mondrians were centred on the four potential locations of the target stimuli. Each Mondrian consisted of sinusoidal square gratings (d = 0.6°), which changed their colour and position randomly at a frequency of 20 Hz. Each grating's texture was shifted every 16.6ms (i.e. each frame of the monitor with 60Hz refresh rate) to generate apparent motion. Visual stimuli were presented at four possible locations that were equidistant from a central fixation spot. They were framed by a grey aperture (thickness: 0.15°, luminance: 110 cd/m^2) of 8.97° x 14.15° in diameter to aid binocular fusion. Mask and target screen allocation (right, left eye) alternated between eyes across trials, to enhance suppression.

Auditory stimuli were 50ms bursts of white noise. They were presented via six external speakers, placed above and below the monitor at 64 dB sound pressure level. Upper and lower speakers were aligned vertically and located centrally, 3° to the left and 3° to the right of the monitor's centre (i.e. aligned with the flash location along the azimuth).

Psychophysical stimuli were generated and presented on a PC running Windows XP using the Psychtoolbox version 3.0.11 (Brainard, 1997) running on MATLAB R2014a (Mathworks, Natick, Massachusetts). Staircase procedures were implemented using the Palamedes toolbox (Kingdom & Prins, 2010).

Visual stimuli were presented dichoptically using a gamma-corrected 30” LCD monitor with a resolution of 2560 x 1600 pixels at a frame rate of 60 Hz (NVIDIA Quadro 600 graphics card). Auditory stimuli were digitized at a sampling rate of 44.8 kHz via an M-Audio Delta 1010LT sound card. Exact audiovisual onset timing was confirmed by recording visual and auditory signals concurrently with a photo-diode and a microphone.

Experiment 1: Design

In a spatial ventriloquist paradigm, participants were presented with an auditory burst of white noise emanating from one of three potential locations: left, centre or right. In synchrony with the sound, one eye was presented with (i) no flash or a brief flash in participants’ (ii) left or (iii) right hemifield under dynamic continuous flash suppression to the other eye (Maruya et al., 2008). Hence, the 3 x 3 factorial design manipulated (1) ‘flash’ (3 levels: left flash, right flash, no flash) and (2) ‘sound location’ (3 levels: left sound, central sound and right sound) (Figure 3.1A). In order to enable a flash localization task that is orthogonal to the sound localization, the flash could be presented either in the upper or lower hemifield (i.e. ± 1.2° elevation from a grey central fixation dot). Hence, the flash was presented in the upper left quadrant, lower left quadrant, upper right quadrant or lower right quadrant (n.b. visual localization is highly precise close to the fixation point and has been shown to be equivalent for spatial discrimination along elevation and azimuth (Dobreva, O’Neill, & Paige, 2012)).

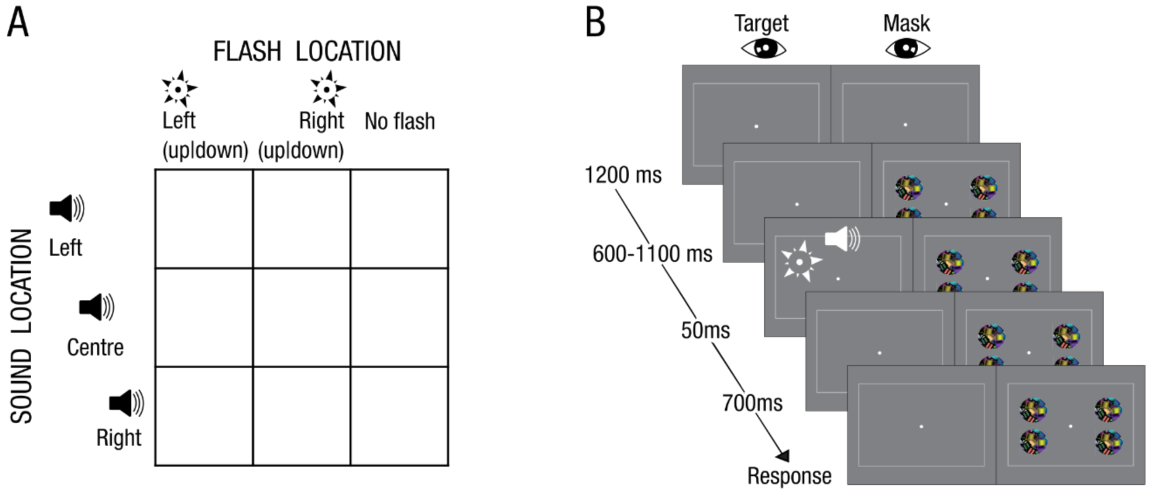

Figure 3.1 Experimental paradigm and procedure. A. Experimental design: 3 x 3 factorial design with the factors: (1) Flash location: left (up|down), right (up|down), no flash; (2) Sound location: left, centre, right. The trials were categorized according to participants' subjective visibility: Clear Image, Almost Clear Image, Weak Glimpse, Not Seen. B. Example trial and procedure of dynamic flash suppression.

Each trial started with the presentation of the fixation dot for a duration of 1200ms (Figure 3.1B). Next, participants were presented with dynamic Mondrians to one eye that suppressed their awareness of signals presented to the other eye (dynamic continuous flash suppression). After a random interval of 600-1100ms, a sound was played from one of three potential locations. On the flash present trials, a white disc was presented in one of the four quadrants for 50ms in synchrony with the sound. The Mondrian masks were presented on the screen until participants had responded to all questions.

On each trial, participants responded to three questions in a self-paced manner within a total response window of 5s: First, they reported the location of the beep (left, centre, right) via a three choice key press. Second, they rated the visibility of the flash (clear image, almost clear image, weak glimpse, not seen) according to a previously published Perceptual Awareness Scale (Ramsøy & Overgaard, 2004; Sandberg, Timmermans, Overgaard, & Cleeremans, 2010)

(PAS) via a four choice key press. This Perceptual Awareness Scale encouraged participants to categorize trials as invisible, only if they were 'completely invisible'. Third, they reported the location of the flash (upper or lower hemifield) via a two choice key press. Critically, we designed orthogonal auditory and visual tasks to minimize decisional biases between visual and auditory localization responses. In order to minimize response interference between responding to the set of three questions, we ensured that the responses mapped to distinct sets of buttons (i.e. 9 different buttons in total). The button/hand assignment and order of questions was counterbalanced across participants (for detailed keyboard mapping please see Figure A.1 in the Appendix A).

This visibility judgment provided a subjective awareness criterion. Critically, prior to the main experiment we adjusted the flash's luminance in the adaptive staircases individually for each participant, such that the flash was visible only on 40% of the trials. This allowed us to quantify multisensory interactions as indexed by spatial ventriloquism (i.e. audiovisual spatial bias) for flashes that were visible (i.e. pooled over clear image, almost clear image, weak glimpse) or invisible (i.e. subjective awareness criterion (Dehaene & Changeux, 2011)). Further, we could assess the information that is available for visual spatial localization during invisible trials and select participants that were not better than chance when locating flashes that they judged as invisible (i.e. the so-called chance performers). The latter allowed us to investigate the influence of flashes on sound localization, when they were invisible and unaware in a so-called objective sense (i.e. objective awareness criterion (Dehaene & Changeux, 2011)).

Prior to the main experiment, participants were familiarized with stimuli and task. In particular, we adjusted the flash luminance in the adaptive staircases (step size up: 8.8 cd/m^2, step size down: 13.2 cd/m^2), such that the flash was visible on 40% of the trials. The adaptive staircases were applied using a slightly modified experimental paradigm where the sound was presented always from the middle, the flash in one of the four quadrants and participants reported only flash visibility (yes, no) and location (up, down). After an initial long staircase (min 200 trials), we performed four times two interleaved adaptive staircases (convergence criterion: 8 reversals within last 10 trials).

During the main experiment, participants completed a total of 8 experimental sessions, resulting in a total of 432 trials (i.e. 64 trials for each flash present condition and 16 trials for each flash absent condition). To maintain the targeted proportion of invisible trials (i.e. 40% visible trials), a staircase procedure was also used throughout the main experiment. To minimize the variability of the flash luminance during the main experiment, we adjusted brightness of the flash in smaller step sizes (3.3 cd/m^2) and only after 4 consecutive 'not seen' responses or after 3 consecutive 'seen' (including all three "partially visible" levels: clear, almost clear & weak glimpse) responses.

Design limitations of Experiment 1 and motivation for Experiment 2

In the first study, flash luminance was adjusted throughout the experiment to maintain a visibility level of approximately 40 % (i.e. 60 % of the trials were judged as invisible based on the four level Perceptual Awareness Scale (Ramsøy & Overgaard, 2004; Sandberg et al., 2010)). This approach is ideal to ensure an approximate visibility level of 40% across all participants. However, it raises the possibility that the ventriloquist effect may be driven by flash stimuli with higher luminance values.

In the second study, we therefore adjusted the flash luminance only during the initial staircases individually for each subject, but held it constant throughout the entire main experiment. This experimental choice ensured that we compare the effect of physically identical flashes that were judged as visible or invisible on sound perception. Yet, because the subjective flash visibility fluctuates throughout the main experiment, this experimental choice induces significant variability in number of invisible and visible trials across participants. To ensure comparable reliability of parameter estimates across participants, we excluded subjects with insufficient number of trials (see exclusion criteria below). Yet, we note that the results were basically equivalent if all participants were included.

Experiment 2: Design

The second study was identical to the first except that the flash luminance was adjusted only prior to the main experiment for each participant, but kept constant throughout the main experiment.

Analysis for experiments 1 and 2

For data analysis, we reduced the four visibility levels to two visibility levels: 1. Visible = clear image + almost clear image + weak glimpse and 2. Invisible = not seen. Further, we pooled over flashes in upper and lower fields given their negligible elevation (i.e. ± 1.2° visual angle). Unless otherwise stated, statistical analysis was identical for the two experiments.

For each participant, we coded their sound location responses as -1 for left, 0 for centre and 1 for right across trials. We estimated the *perceived sound location* for each of the 2 (flash location: left, right) x 3 (sound location: left, centre, right) conditions by averaging the localization responses across trials. Next, we averaged the perceived sound locations separately for trials where the flash was presented on the left and right and computed the difference in average perceived sound location for 'visual right' minus 'visual left' trials as the index for the spatial ventriloquist effect (for illustration please see Figure A.2 in Appendix A). A positive value of this index indicates that subject's perceived sound location shifted towards the visual stimulus location (i.e. 'attraction') and a negative value indicates that it is shifted away from the visual stimulus location (i.e. repulsion). A ventriloquist effect of zero means that participants were not influenced consistently across trials by the location of the flash.

This difference in perceived sound location, i.e. the ventriloquist effect, was then used as the dependent variable for all subsequent analyses. If the visual signal location attracts the perceived sound location, we would expect the difference to be significantly greater than zero. Given our a priori directed hypothesis, all p-values are reported for right-tailed one sample t-tests.

The data analysis was performed in MATLAB (Mathworks, Natick, Massachusetts) with exception of unidirectional Bayes factors, which were computed using JASP (Marsman & Wagenmakers, 2017).

Exclusion criteria

For the reported results, we limited the analysis to subjects based on the following two exclusion criteria: First, we included only subjects that provided reliable visibility judgments as indicated by 'better than chance localization accuracy' (based on a binomial test) for visible flashes (i.e. exclusion of six subjects from experiment 1 and three subjects from experiment 2). Second, to ensure reliable parameter estimation, we included only those participants who had at least 10 trials in each of the 2 (flash location: left vs. right) x 3 (sound location: left, middle, right) conditions, for both visible and invisible categories respectively (i.e. exclusion of three subjects from experiment 1 and six subjects from experiment 2). For individual distribution of PAS ratings, see Figure A.3 in Appendix A. These exclusion criteria ensured that the computation of the ventriloquist effect was based on at least 60 trials. Yet, the minimal number of trials was even higher and amounted to 106 trials.

Further, we would like to emphasize that the results were basically equivalent when including all participants (apart from one for whom the ventriloquist effect could not be computed for invisible trials because only two trials were categorized as invisible and they did not fall into the corresponding conditions). In other words, significant results were again significant, non-significant results again non-significant, when no participants were excluded.

Direct comparison of spatial ventriloquism for visible and invisible trials

We investigated whether the audiovisual spatial bias (i.e. ventriloquist effect) was significantly different for visible (i.e. clear image, almost clear image and weak glimpse) and invisible trials using paired t-tests.

Spatial ventriloquism for visible and invisible trials

We investigated whether the ventriloquist effect was present independently for both invisible and visible flashes. Hence, we computed the ventriloquist effect separately for trials where visual signals were judged visible or invisible and tested whether the ventriloquist effect was significantly greater than zero in right-tailed one sample t-tests independently for visible and invisible trials. Demonstrating a ventriloquist effect for invisible trials suggests that flashes can influence perceived sound location, when participants are subjectively not aware of them (i.e. subjective awareness criterion (Dehaene & Changeux, 2011)). Further, we investigated whether participants at the group level were at chance when locating a flash they judged invisible by comparing their accuracy scores against 50% chance performance in a right tailed one-sample t-test. For p values greater than 0.05, we computed Bayes factors to provide further evidence for the null hypothesis (i.e. 50% chance performance).

Spatial ventriloquism for invisible trials in chance performers

We asked whether invisible flashes are able to influence the perceived sound location, even in participants that are not better than chance when locating flashes they judged as invisible (i.e. objective awareness criterion (Dehaene & Changeux, 2011)). Using directionally informed tests in i. classical statistics and ii. Bayesian inference we identified chance performers based on a binomial test on their flash localization performance on trials in which the flash was judged as invisible. First, to link with previous reports in the literature, we defined chance performers individually based on a 'null-result' using a directional binomial test (i.e. 'not significantly better than chance') in classical statistics (Bahrami et al., 2010). Second, as a 'null-result' in classical statistics is not decisive, we also used Bayesian statistics that allows one to quantify and compare the evidence for the null model that embodies the null-hypothesis in relation to an alternative model. Hence, using Bayes factors we compared a binomial distribution model that a priori fixes the probability to 0.5 (i.e. null model of chance performance) with one that includes the probability parameter p as a free parameter constrained by a positive prior distribution (i.e. directional binomial test). Please note that imposing a positive prior distribution makes the Bayesian test more stringent for defining chance performers.

Nevertheless, because both selection criteria were applied in a non-crossvalidated fashion as is currently common in the field (Z. Chen & Saunders, 2016; Eo, Cha, Chong, & Kang, 2016; Faivre et al., 2014; Sklar et al., 2012), the definition of so-called chance performers can at least in part be susceptible to noise (inter-trial variability). As has been explained in detail in Shanks (Shanks, 2017), the flash localization accuracy may be lower in chance than non-chance performers in this experimental session because of performance noise and hence if measured again may increase, a statistical phenomenon referred to as 'regression towards the mean'. Conversely, participants' sound localization performance may be affected by - partly - independent noise. As a result, we may underestimate and falsely define observers as chance performers and conversely overestimate the 'unaware' ventriloquist effect.

Influence of question order on flash localization accuracy and ventriloquism

As described earlier, we counterbalanced the task order (either: flash location – visibility – sound location or: sound location – visibility – flash location) across participants, because the task order can affect our analysis results in three important ways:

First, the presentation order of the questions can influence observer's performance accuracy on the different tasks. For instance, as a result of memory noise, observers' flash localization accuracy may be reduced when the flash localization task was presented last with additional consequences on the size of the ventriloquist effect. We therefore compared flash localization accuracy and the size of spatial ventriloquism for first vs. last task in a two-sample t-test.

Second, as a result of the reduced flash localization accuracy, we may have classified participants as chance performers mainly when the flash localization task was presented last. To assess the effect of task order on the classification as chance performers, we compared the number of chance performers across the groups, where the flash localization task was presented first vs. last in a Chi-square test.

Third, to assess the effect of task order on the exclusion of participants we compared the number of excluded participants across the groups where the flash localization task was presented first vs. last in a Chi-square test.

Results (Experiments 1 and 2)

Direct comparison of spatial ventriloquism for visible and invisible trials

A paired t-test comparing the ventriloquist effect for invisible and visible trials demonstrated a significantly greater ventriloquist effect for visible than invisible flashes (experiment 1: all participants: mean difference ± SEM = 0.44 ± 0.06, t(32) = 7.99, p < 0.001; chance performers: 0.47 ± 0.05, t(27) = 8.57, p < 0.001; experiment 2: all participants: 0.46 ± 0.07, t(17) = 6.45, p < 0.001; chance performers: 0.46 ± 0.07, t(12) = 6.34, p < 0.001).

Spatial ventriloquism for visible and invisible trials

To investigate whether the ventriloquist effect emerged for visible and invisible flashes we performed one sample t-tests independently for each of these two visibility levels.

For visible trials, we observed a significant ventriloquist effect (i.e. audiovisual spatial bias) as expected based on numerous previous studies (Bertelson & Radeau, 1981; Wallace et al., 2004) (experiment 1: mean ± SEM = 0.52 ± 0.06, right-tailed t(32) = 8.62, p < 0.001; Cohen's d

± 95% Confidence Interval: 1.5 ± 0.55 ; experiment 2: 0.51 ± 0.06, t(17) = 9.14, $p < 0.001$; Cohen's d: 2.15 ± 0.82).

Crucially, we also observed a significant ventriloquist effect for flash stimuli that participants judged as invisible (experiment 1: 0.08 ± 0.02, t(32) = 3.86, $p < 0.001$; Cohen's d: 0.67 ± 0.5; experiment 2: 0.09 ± 0.03, t(17) = 3.22, p = 0.003; Cohen's d: 0.76 ± 0.68). Participants' flash localization accuracy was slightly above chance (experiment 1: mean ± SEM = 51.3 ± 0.7%; right-tailed t-test against 50% chance performance: t(32) = 1.88, p = 0.034; corresponding BF_{01} = 1.72; experiment 2: 53.6 ± 2%, t(17) = 0.91 , p = 0.047; BF_{01} = 1.69), therefore the results reported so far only provide evidence that flashes that participants consider invisible (i.e. subjective awareness criterion) are still able to elicit a robust ventriloquist effect.

Spatial ventriloquism for invisible trials in chance performers defined based on classical statistics

To ensure that the ventriloquist effect for invisible trials was not driven by participants that had residual visual information for visual flash localization, we repeated this analysis using the more stringent so-called objective criterion for perceptual awareness. Hence, we included only those subjects that were individually not better than chance when locating an invisible flash based on binomial testing (i.e. objective awareness criterion). The constraint of individual chance performance reduced the number of subjects that could be included in the analysis (experiment 1: n = 28; experiment 2: n = 13).

Nevertheless, despite the reduced number of subjects, we still observed a highly significant ventriloquist effect for invisible trials (experiment 1: 0.07 ± 0.02, t(27) = 3.22, p = 0.002;

Cohen’s d : 0.61 ± 0.54; experiment 2: 0.08 ± 0.03, t(12) = 2.84, p = 0.007; Cohen’s d : 0.79 ± 0.8) (Figure 3.2).

In other words, both experiments jointly demonstrated that an invisible flash attracted the perceived sound location even in subjects that were not better than chance when locating the flash, they judged as invisible. The flash localization accuracy (i.e. across subjects mean) at the group level was not significantly better than chance but nearly equal to 50% (experiment 1: 50.3 ± 0.6%; t(27) = 0.41, p = 0.341; BF_{01} = 3.53; experiment 2: 49.5 ± 1.1%, t(12) = - 0.5, p = 0.687; BF_{01} = 4.97) (Figure 3.2 A&C). Hence, when selecting participants based on the objective awareness criterion, the Bayes factors at the group level provided strong evidence for the null-hypothesis, i.e. that participants were not better than chance when locating the flash, independently for each experiment.

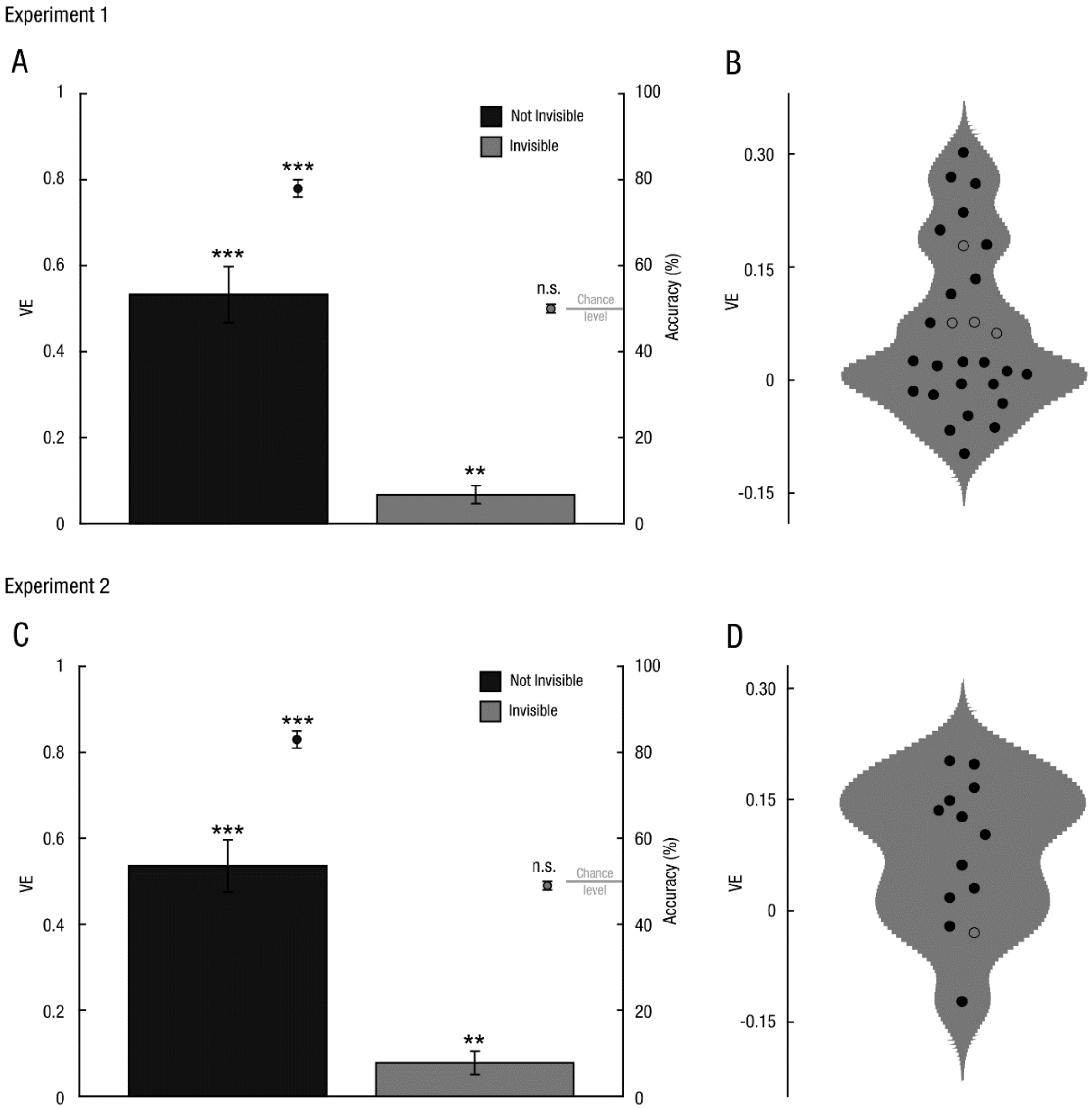

Figure 3.2 VE for chance performers: experiment 1 (n = 28) & experiment 2 (n = 13). A & C Bar plots showing the ventriloquist effect in chance performers (VE, across subjects mean ± SEM) for visible and invisible flashes (left axis). The VE was significantly greater than zero for both visible and invisible trials. The markers show the accuracy (across subjects mean ± SEM) for flash localization (right axis: percentage correct). B & D Violin plots showing the distribution of individual ventriloquist effects for invisible trials in chance performers identified based on classical and Bayesian binomial tests. All dots represent subjects with not significantly better than chance performance based on classical statistics. Filled dots show subjects, for which BF01 for Bayesian binomial test was also greater than 3 (i.e. positive evidence for the null model of chance performance). The mass of the probability distribution is clearly above zero. Markers show the individual data points. *** $p < 0.001$, ** $p < 0.01$, n.s. $p > 0.05$.

Spatial ventriloquism for invisible trials in chance performers individually defined based on Bayesian statistics

As classical statistics does not allow the acceptance of the null-hypothesis, we also used unidirectional Bayesian tests that can formally provide positive evidence for the null-hypothesis of chance performance individually for each participant. Specifically, we computed Bayes factors comparing the evidence for the null-model of chance performance with the evidence for alternative model of better than chance performance and selected only subjects as chance performers with Bayes factors > 3 (i.e. positive evidence for the null-model). This reduced the number of included chance performers in both experiments (experiment 1: n = 24, experiment 2: n = 12, mean flash localization accuracy: 49.5 ± 0.6% and 48.9 ± 1% respectively).

Nevertheless, despite this more stringent objective awareness criterion we again observed a significant ventriloquist effect for invisible trials (experiment 1: 0.06 ± 0.02, right-tailed t(23) = 2.61, p = 0.008; Cohen's d: 0.53 ± 0.58; experiment 2: 0.09 ± 0.03, t(11) = 3.09, p = 0.005; Cohen's d: 0.89 ± 0.84).

Influence of question order on flash localization accuracy and ventriloquism

A two sample t-test did not reveal a significant effect of question order on flash localization accuracy (experiment 1: visible: p = 0.209, t(31) = -1.28; invisible: p = 0.454, t(31) = -0.76 ; experiment 2: visible: p = 0.822, t(16) = 0.23, invisible: p = 0.176, t(16) = -1.42) or the size of the ventriloquist effect (experiment 1: visible: p = 0.703, t(31) = -0.39, invisible: p = 0.141, t(31) = -1.51; experiment 2: visible: p = 0.537, t(16) = -0.63, invisible: p = 0.523, t(16) = -0.65).

Likewise, a Chi-square test did not reveal a significant effect of question order on the number of subjects classified as chance performer using criteria based on Binomial testing (experiment 1: $\chi^2 = 0.50$, $p = 0.478$; experiment 2: $\chi^2 = 1.68$, $p = 0.196$) or Bayesian statistics (experiment 1: $\chi^2 = 2.25$, $p = 0.134$; experiment 2: $\chi^2 = 0.45$, $p = 0.502$). Further, question order did not significantly affect the exclusion of participants (experiment 1: $\chi^2 = 0.90$, $p = 0.343$; experiment 2: $\chi^2 = 0.37$, $p = 0.541$).

Discussion

Using continuous flash suppression and spatial ventriloquism, we demonstrate that unconscious signals in the visual modality influence how humans construct their auditory perceptual world. In particular, we have shown that flashes judged as invisible alter the perceived location of concurrent sounds, even when participants are at chance when locating the flash. These results suggest that auditory and visual inputs are integrated into spatial representations at least to some extent in the absence of subjective and objective perceptual awareness.

Accumulating evidence has shown that audio-visual integration of speech information is abolished when visual facial movements are rendered unconscious via multistable perception, binocular rivalry or flash suppression (Munhall et al., 2009; Palmer & Ramsey, 2012) highlighting the role of perceptual awareness in multisensory integration. This raises the question whether consciousness is a generic prerequisite for multisensory integration and is also required for or associated with interactions of spatial signals as indexed by the ventriloquist effect.

Our findings demonstrate that spatial ventriloquism is profoundly modulated by the visibility of the flash. While a strong ventriloquist effect was observed for visible trials, it was attenuated when the flash was judged as invisible. Nevertheless, a robust ventriloquist effect was observed across both experiments for trials when participants judged the flash as invisible (i.e. subjective awareness criterion).

Moreover, across both experiments, the ventriloquist effect persisted even for invisible flashes when participants showed chance performance on flash localization (i.e. objective awareness criterion).

Collectively, our two experiments show that invisible flashes, that human observers are not aware of, can influence where they report sounds, that they are aware of.

Invisible flashes may influence sound localization during continuous flash suppression via at least three distinct neural circuitries. First, an invisible flash may interact with auditory signals via subcortical mechanisms such as the colliculo-pulvinar pathway (Cappe, Morel, Barone, & Rouiller, 2009; Hackett et al., 2007) that has previously been implicated in mediating activations along the dorsal stream into the intraparietal sulcus under CFS (Fang & He, 2005), but see (Ludwig & Hesselmann, 2015; Ludwig, Kathmann, Sterzer, & Hesselmann, 2015; Rothkirch & Hesselmann, 2018). Because participants were engaged in a spatial localization task and the ventriloquist effect relies on integration of spatial representations from vision and audition, the dorsal stream may be critical in our paradigm (Rohe & Noppeney, 2015b, 2016, 2018). Second, it may modulate sound processing via sparse direct connectivity between primary auditory and visual areas (Cappe & Barone, 2005; Falchier et al., 2002). Third, some flash-induced neural activity may evade flash suppression and propagate across the cortical hierarchy into higher order association areas such as intraparietal sulcus or even prefrontal cortices (Bonath et al., 2014; Dahl, Logothetis, & Kayser, 2009; Falchier et al., 2002; Ghazanfar & Schroeder, 2006; Macaluso & Driver, 2005; Rohe & Noppeney, 2015a, 2016; Werner & Noppeney, 2010). While this activation may not be sufficient to allow better than chance flash location, it enables to bias participants' sound localization.

The ventriloquist effect may be smaller for invisible than visible flashes, because invisible flashes may evoke weaker or less reliable activations than visible flashes already at the primary cortical level as a result of state-dependent effects or various sources of internal neural noise (Faisal, Selen, & Wolpert, 2008). The level of neural activity then concurrently determines (i) whether the flash is able to enter perceptual awareness and (ii) the precision of the spatial representation and thereby the strength of the ventriloquist effect (Ma, Beck, Latham, & Pouget, 2006; Yuval-Greenberg & Heeger, 2013). Thus, visible flashes would induce a ventriloquist effect via the same neural circuitries as invisible flashes and induce a greater ventriloquist effect, as they induce higher neural activity and thus more precise spatial representations in visual cortices.

Alternatively, visible flashes may induce a stronger ventriloquist effect by employing additional neural circuitries (e.g. via higher order association areas) that are not engaged by weaker invisible flashes. In this account, the spatial representations elicited by a flash at the primary cortical level may be preserved, yet be less effective in influencing the sound processing system. This latter account dovetails nicely with current perspectives on the neural organization of multisensory integration. Specifically, auditory and visual information are thought to be integrated via multiple circuitries including subcortical mechanisms, direct connectivity between primary sensory areas and convergence in higher order association areas (Ghazanfar & Schroeder, 2006; Macaluso & Driver, 2005; Rohe & Noppeney, 2016; Werner & Noppeney, 2010). Moreover, it is well established that multisensory integration progressively increases along the cortical hierarchy with only about 15% neurons showing multisensory properties in primary sensory areas (Bizley et al., 2007) and more than 50% in classical association areas such as intraparietal or superior temporal sulci (Dahl et al., 2009).

Thus, when a visual flash escapes the continuous flash suppression and enters participants' awareness, a strong ventriloquist effect emerges most likely via integration in association areas such as intraparietal sulci (IPS) that contain exuberant multisensory neurons and may potentially amplify multisensory integration via feed-back loops with lower level sensory areas. By contrast, when continuous flash suppression blocks neural activity at least to some extent from propagating into higher order association areas, audio-visual interactions are greatly attenuated or even abolished leading to a smaller ventriloquist effect. Under this 'multiple neural circuitries' account, auditory and visual signals interact most likely at both pre- and post-aware processing stages by placing different demands on distinct neural circuitries (e.g. direct connectivity vs. higher order association cortices).

The combination of different psychophysical blinding methods that affect visual processing at variable depths (Breitmeyer, 2015; Moors, Hesselmann, Wagemans, & van Ee, 2017; Yuval-Greenberg & Heeger, 2013) may enable us to better dissociate between these different mechanisms. For instance, while flash suppression is thought to affect processing in primary visual areas alike contrast modulation (Yuval-Greenberg & Heeger, 2013), attentional blink may alter processing mainly at higher attentional levels. In fact, we suspect that our current paradigm potentially combines both mechanisms by placing attentional demands at four locations.

In conclusion, to our knowledge, our findings provide the first demonstration that invisible flashes can alter and bias where we perceive sounds. These results suggest that low level sensory information can interact across sensory modalities at least to some extent prior to perceptual awareness. Nevertheless, audiovisual interactions as indexed by spatial ventriloquism were stronger for visible relative to invisible flashes that participants were not able to locate better than chance. This raises the possibility that aware visual signals may also engage multisensory mechanisms in higher order association areas or other neural circuitries that are less engaged in the absence of perceptual awareness. Future studies using EEG and fMRI are needed to identify the neural systems that enable audio-visual interactions in the presence and absence of subjective and objective awareness.

Appendix A

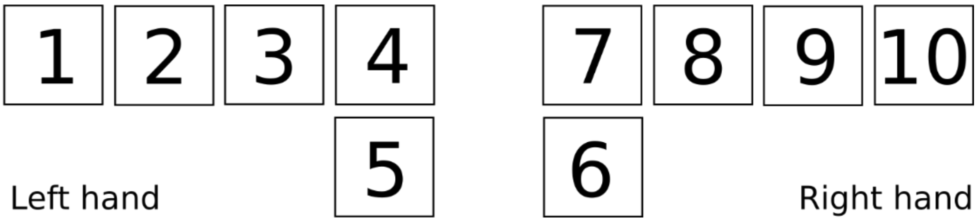

Figure A.1 Keyboard mappings. Participants used 9 different buttons to respond. The button/hand assignment and order of questions was counterbalanced across participants (i.e. subject reported visibility with left hand and flash/sound locations with right hand or vice versa). Version A: visibility (PAS scale) - buttons 1-4; sound location (left, center, right) - buttons 8-10; flash location (top, down) - buttons 6 & 7. Version B: visibility (PAS scale) - buttons 7-10; sound location (left, center, right) - buttons 1-3; flash location (top, down) - buttons 4 & 5.

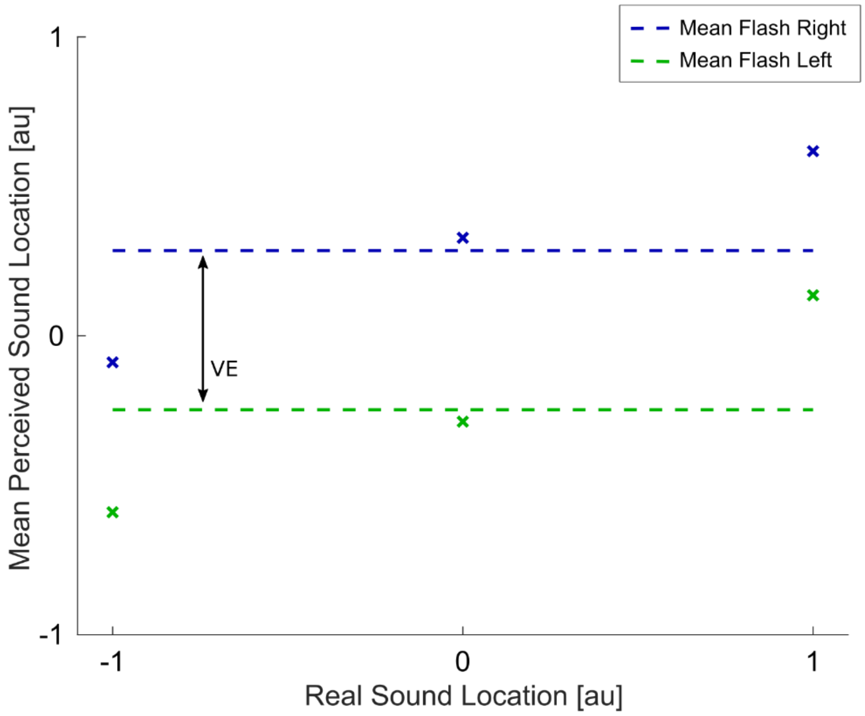

Figure A.2 Computation of Ventriloquist Effect. The crosses indicate the mean perceived sound location (in arbitrary units) for each of the 2 flash (left, right) x 3 sound (left, centre, right) conditions. Mean of the differences between 'visual right' (blue) and 'visual left' conditions serves as index of spatial ventriloquism. Results for visible condition for chance performers from experiment 1 were used for illustration.

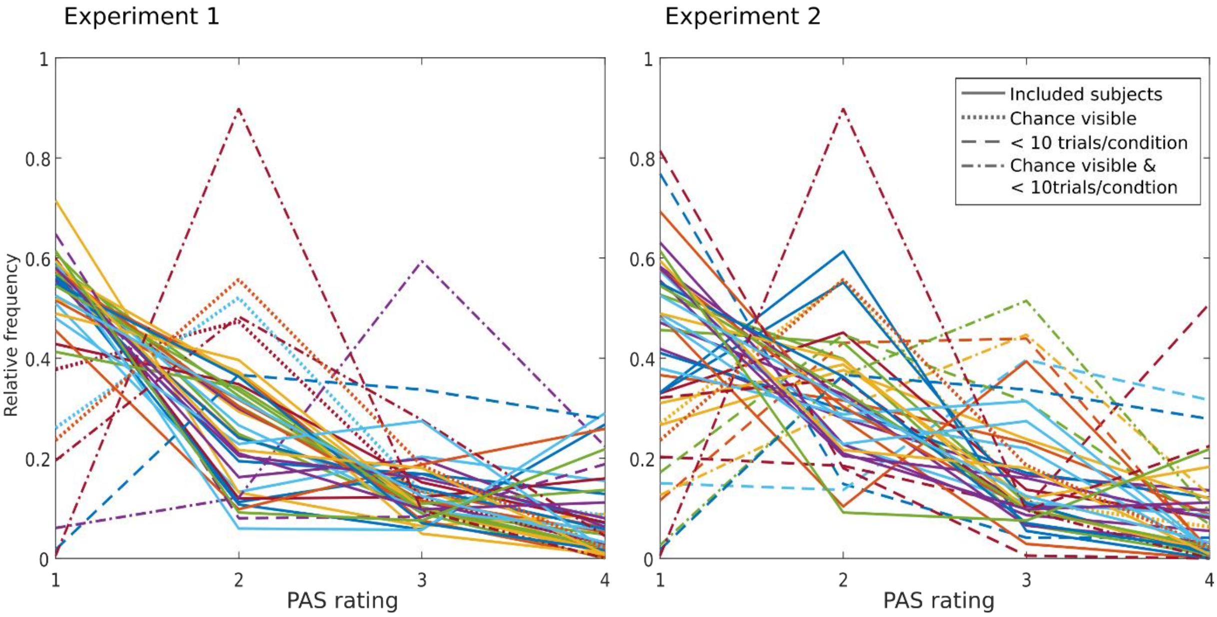

Figure A.3 Visibility judgement. Figures show proportions of perceptual awareness scale (PAS) ratings for subjects from experiment 1 and 2. Each line represents an individual participant. Solid = included in the analysis; dotted = excluded because of chance performance on flash localization for visible trials; dashed = excluded because less than 10 trials in each condition; dotted-dashed = excluded because of both, i.e. chance performance for flash localization accuracy for visible trials and less than 10 trials in each condition.

CHAPTER 4. NEURAL BASIS OF INVISIBLE VENTRILOQUISM

Patrycja Delong, Uta Noppeney

Contributions:

PD and UN designed the study. PD acquired the data. PD analysed the data under supervision of UN. PD wrote the first version of the manuscript.

Results of this chapter were presented at:

The 22nd meeting of the Association for the Scientific Study of Consciousness, Krakow, Poland, 2018

Talk: “The invisible ventriloquist: audio-visual integration in the absence of perceptual awareness”

Abstract

Information integration is considered a hallmark of human consciousness (Baars, 2005). Yet, recent psychophysics work has provided initial evidence for integration of unconscious signals across the senses (Delong et al., 2018). How can information that evades our subjective awareness influence perceptual representation of other senses?

This psychophysics-EEG study combined dynamic continuous flash suppression with spatial ventriloquism - a perceptual illusion where observers perceive a sound shifted towards a synchronous, spatially incongruent flash. Dynamic CFS obliterated visual awareness only on a fraction of trials allowing us to compare the ventriloquist effect for physically identical flashes that were i. visible or invisible and ii. located correctly or incorrectly.

Behaviourally, we observed a robust spatial ventriloquism for visible and invisible flashes. Importantly, even when observers mislocated the invisible flash, they perceived the location of the sound as shifted towards the true flash location.

True flash location (left vs. right) could be successfully decoded from early EEG activity irrespectively of the flash's visibility. Yet, on invisible trials, where observers mislocated the flash, this visual-spatial information rapidly dissipated after 300ms. Multivariate EEG analyses showed that on illusion trials, the brain encoded the sound's location biased towards the flash location irrespectively of the flash's visibility and correct flash localization. Collectively, these findings suggest that invisible flashes influence the neural representations of sounds in the absence of subjective awareness via early multisensory interactions. As a consequence, a flash that evades our subjective awareness and that we cannot correctly locate can influence where we perceive sounds.

Our findings unravel the neural mechanisms that enable integration of auditory and visual signals into perceptual spatial representations in the absence of subjective awareness. To our knowledge they are the first compelling demonstration that unconscious signals in one sensory modality can alter the neural and perceptual representations in another sensory modality. A subliminal flash can change where we perceive a supraliminal sound via early multisensory interactions.

Our results challenge current models of consciousness. They demonstrate that consciousness is not a generic prerequisite for information integration, not even in situations where the signals stem from different sensory channels.

Introduction

Information integration is widely regarded as a hallmark of human consciousness (Baars, 2002, 2005; Dehaene & Changeux, 2011). Most notably, the Global Workspace Theory (GWT) assumes that nonconscious processing is confined to local neural circuitries. Only conscious information enters the global workspace and can be broadcast through long range connections across distant brain regions. The GWT therefore predicts that only consciously perceived signals in one sensory modality influence neural processing in other sensory systems. Multiple studies have shown that supraliminal stimuli from one sensory modality can influence neural processing of subliminal stimuli from another modality and thereby boost them into perceptual awareness (Adam & Noppeney, 2014; Aller et al., 2015; Alsius & Munhall, 2013; Cox & Hong, 2015; Lunghi & Alais, 2015; Lunghi et al., 2010, 2017; Ngo & Spence, 2010; Olivers & Van der Burg, 2008; Salomon et al., 2017, 2015; Zhou et al., 2010). These directed influences from supraliminal to subliminal processing are in line with GWT, because the neural activity associated with processing of a conscious signal can enter the global network and thereby impact local processing of subliminal signals in other sensory modalities. Yet, recent psychophysics work has also provided initial evidence for cross-modal associative learning between two unconscious signals (Faivre et al., 2014; Scott, Samaha, Chrisley, & Dienes, 2018) or influences of an unconscious visual signal on conscious sound perception (Delong et al., 2018). In the latter case, an invisible flash that was presented in synchrony yet displaced from a sound was shown to produce the ventriloquist illusion: shift (i.e. bias) of observers' perceived sound location towards the flash (Bertelson & Aschersleben, 1998). These influences of unconscious signals on signal processing in other sensory systems are difficult to reconcile with the GWT, because the neural activity associated with unconscious processing should be confined to local neural circuitries. They raise the critical question of how information that evades our awareness in one sensory modality can impact neural and perceptual representations in other sensory systems. How do invisible flashes influence where we perceive sounds?

Critically, unconscious processing can be defined based on subjective or objective awareness thresholds with their strengths and limitations (Dehaene & Changeux, 2011). Subjective thresholds reflect observers' phenomenal experience, but are susceptible to criterion shifts driven by observers' confidence (Björkman, Juslin, & Winman, 1993). Conversely, objective thresholds have been criticised for being too strict, focusing on unconscious processing of degraded stimuli (Lau & Passingham, 2006). Moreover, the common methodological approach of post-hoc selection of participants that fulfil the objective criterion of chance performance involves serious statistical problems of regression towards the mean (see: Shanks, 2017). Moreover, it has recently been argued that only subjective awareness thresholds are appropriate to test global workspace predictions, as i. neural processing of subliminal stimuli in local neural circuitries can enable better than chance performance thereby violating the objective awareness criteria and ii. stimuli that entered the global workspace should be available for visibility report (Scott et al., 2018).

The present EEG study combined dynamic Continuous Flash Suppression (Maruya et al., 2008; Tsuchiya & Koch, 2005) with spatial ventriloquism - a perceptual phenomenon in which the perceived location of a sound shifts towards the location of a synchronous, but spatially incongruent flash. On each trial, observers reported their perceived sound and flash locations and rated the flash's visibility. Using psychophysics and EEG multivariate pattern analysis, we resolved the neural dynamics of spatial information for visible and invisible flashes across time and assessed how visual information impacts auditory spatial processing as indexed by neural and behavioural ventriloquism. Crucially, dynamic flash suppression obliterated visual awareness only on a fraction of trials allowing us to compare behavioural and neural spatial ventriloquism for physically identical flashes that were i. visible or invisible and ii. located correctly or incorrectly.

Methods

Participants

After giving informed consent, 103 (78 females, 8 left-handed, mean age: 21.5 years, standard deviation: 4.9, range: 18-41; 41 participants were included in (Delong et al., 2018)) healthy young adults took part in the psychophysics experiment; 72 of those participants were included in the analysis (see inclusion criterion section). Eighteen of those subjects participated (13 females, 2 left-handed, mean age: 21.2 years, standard deviation: 4.2, range:

18-31) in the subsequent EEG experiment. The study was performed in accordance with the principles outlined in the Declaration of Helsinki and was approved by the local ethics review board of the University of Birmingham.

To determine sample size for the EEG experiment we used effect size (Cohen's d ≈ 0.7) for ventriloquist effect for invisible trials based on the psychophysics studies described in Delong et al. 2018. For statistical power of 0.9 we obtained n = 18. We continued acquiring subjects for the psychophysics experiment until the number of EEG data sets was equal to required sample size (i.e. excluded subjects were replaced; see section inclusion criteria).

Stimuli and apparatus

Participants sat in a dimly lit room in front of a computer monitor at a viewing distance of 95cm. They viewed one half of the monitor with each eye using a custom-built mirror stereoscope. Visual stimuli were composed of targets and masks that were presented on a grey, uniform background with a mean luminance of 15.6 cd/m^2. On the 'flash present' trials, one eye viewed the target stimulus (i.e. the flash), which was a grey disc (Ø 0.3°) presented for 50ms in the upper left, lower left, upper right or lower right quadrant, i.e. at ± 3° visual angle along the azimuth and ± 1.2° elevation from a grey central fixation dot. The elevation of ± 1.2° was selected to enable effective multisensory interactions between flash and sound irrespective of flash elevation. The luminance of the flash was adjusted individually via adaptive staircases to obtain 60% invisible trials. To suppress the flash's perceptual visibility, four dynamic Mondrians (Ø 2.08°, mean luminance: 48 cd/m^2) were shown to the other eye (Maruya et al., 2008). In dynamic CSF, static rectangles (Tsuchiya & Koch, 2005) were replaced with dynamically moving gratings (Aller et al., 2015; Maruya et al., 2008). The Mondrians were centred on the four potential locations of the target stimuli. Each Mondrian consisted of sinusoidal square gratings (d = 0.6°) which changed their colour and position randomly at a frequency of 20Hz. Each grating's texture was shifted every 16.6ms (i.e. each frame of the monitor with 60Hz refresh rate) to generate apparent motion. Visual stimuli were presented at four possible locations that were equidistant from a central fixation spot. They were framed by a grey aperture (thickness: 0.15°, luminance: 110 cd/m^2) of 8.97° x 14.15° in diameter to aid binocular fusion. Mask and target screen allocation (right, left eye) alternated between eyes across trials, to enhance suppression.

Auditory stimuli were 50ms bursts of white noise. They were presented via six external speakers, placed above and below the monitor at 64 dB sound pressure level. Upper and lower speakers were aligned vertically and located centrally, 3° to the left and 3° to the right of the monitor's centre (i.e. aligned with the flash location along the azimuth).

Psychophysical stimuli were generated and presented on a PC running Windows XP using the Psychtoolbox version 3.0.11 (Brainard, 1997) running on MATLAB R2014a (Mathworks, Natick, Massachusetts). Staircase procedures were implemented using Palamedes toolbox (Kingdom & Prins, 2010).

Visual stimuli were presented dichoptically using a gamma-corrected 30" LCD monitor with a resolution of 2560 x 1600 pixels at a frame rate of 60Hz (NVIDIA Quadro 600 graphics card). Auditory stimuli were digitized at a sampling rate of 44.8 kHz via an M-Audio Delta 1010LT sound card. Exact audiovisual onset timing was confirmed by recording visual and auditory signals concurrently with a photodiode and a microphone.

Experimental Design

In a spatial ventriloquist paradigm, participants were presented with an auditory burst of white noise emanating from one of three potential locations: left, centre or right. In synchrony with the sound, one eye was presented with (i) no flash or a brief flash in participants' (ii) left or (iii) right hemifield under dynamic continuous flash suppression to the other eye (Maruya et al., 2008). Hence, the 3 x 3 factorial design manipulated (1) 'flash' (3 levels: left flash, right flash, no flash) and (2) 'sound location' (3 levels: left sound, central sound and right sound) (Figure 4.1A). In order to enable a flash localization task that is orthogonal to the sound localization, the flash could be presented either in the upper or lower hemifield (i.e. ± 1.2° elevation from a grey central fixation dot). Hence, the flash was presented in the upper left quadrant, lower left quadrant, upper right quadrant or lower right quadrant (n.b. visual localization is highly precise close to the fixation point and has been shown to be equivalent for spatial discrimination along elevation and azimuth (Dobreva et al., 2012)).

Each trial started with the presentation of the fixation dot for a duration of 1200ms (Figure 4.1B). Next, participants were presented with dynamic Mondrians to one eye that suppressed their awareness of signals presented to the other eye (dynamic continuous flash suppression). After a random interval of 600-1100ms, a sound was played from one of three potential locations. On the flash present trials, a white disc was presented in one of the four quadrants for 50ms in synchrony with the sound. Response cues were presented 750ms poststimulus. The Mondrian masks were presented on the screen until participants had responded to all questions.

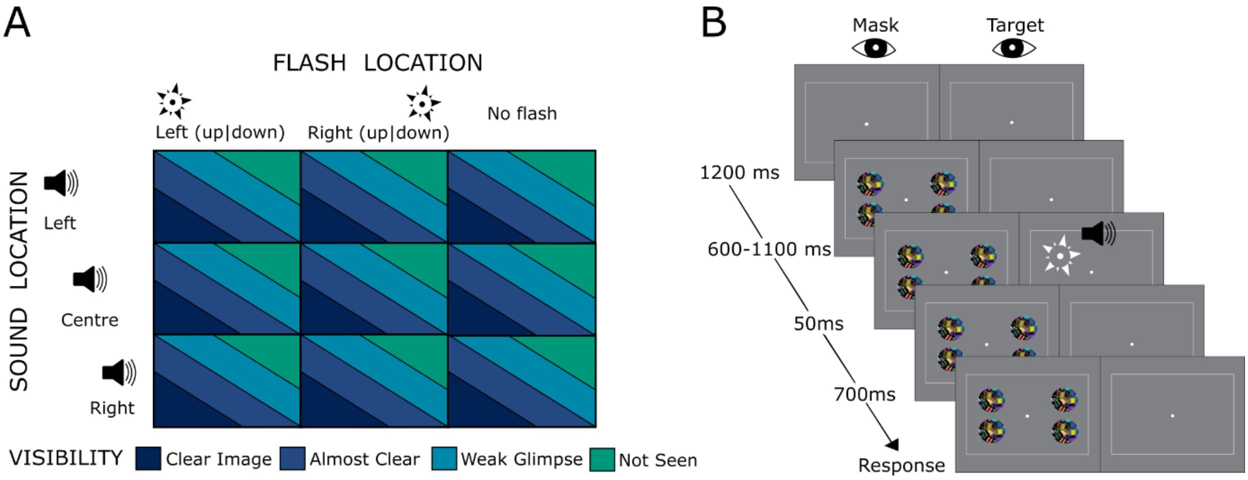

Figure 4.1 Experimental paradigm and procedure. A. Experimental design: 3 × 3 factorial design with the factors: (1) Flash location: left (up|down), right (up|down), no flash; (2) Sound location: left, centre, right. The trials were categorized according to participants' subjective visibility: Clear Image, Almost Clear Image, Weak Glimpse, Not Seen. B. Time course of an example trial.

On each trial, participants responded to three questions in a self-paced manner within a total response window of 5s: first, they reported the location of the beep (left, centre, right) via a three choice key press. Second, they rated the visibility of the flash (clear image, almost clear image, weak glimpse, not seen) according to a previously published Perceptual Awareness Scale (PAS) (Ramsøy & Overgaard, 2004; Sandberg et al., 2010) via a four choice key press. This Perceptual Awareness Scale and experimenter's explicit instructions encouraged participants to categorize trials as invisible, only if they were 'completely invisible'. Third, they reported the location of the flash (upper or lower hemifield) via a two choice key press. Critically, we designed orthogonal auditory and visual tasks to minimize decisional biases between visual and auditory localization responses. In order to minimize response interference between responding to the set of three questions, we ensured that the responses mapped to distinct sets of buttons (i.e. 9 different buttons in total). The button/hand assignment and order of questions was counterbalanced across participants.

This visibility judgment provided a subjective awareness criterion. Critically, we adjusted the flash's luminance in adaptive staircases individually for each participant, such that the flash was visible only on 40% of the trials. This allowed us to quantify multisensory interactions as indexed by spatial ventriloquism (i.e. audiovisual spatial bias) for flashes that were visible (i.e. pooled over 'clear image', 'almost clear image', 'weak glimpse') or invisible (i.e. 'not seen', subjective awareness criterion (Dehaene & Changeux, 2011)).

Experimental procedure

The study included a one day psychophysics experiment and four day EEG experiment for a subset of those subjects.

Prior to all experiments, we adjusted the flash luminance in adaptive staircases (step size up: 8.8 cd/m^2, step size down: 13.2 cd/m^2), such that the flash was visible on 40% of the trials. The adaptive staircases were applied using a slightly modified experimental paradigm where the sound was presented always from the middle, the flash in one of the four quadrants and participants reported only flash visibility (yes, no) and location (up, down). After an initial long staircase (min 200 trials), we performed four times two interleaved adaptive staircases (convergence criterion: 8 reversals within last 10 trials).

The one-day psychophysics experiment included a total of 8 experimental runs, resulting in a total of 432 trials (i.e. 64 trials for each flash present condition and 16 trials for each flash absent condition). In this initial psychophysics study, the flash luminance was adjusted throughout the experiment with adaptive staircases to maintain a visibility level of approximately 40 % (i.e. 60 % of the trials were judged as not seen based on the four level Perceptual Awareness Scale (Ramsøy & Overgaard, 2004; Sandberg et al., 2010)). To minimize the variability of the flash luminance during the psychophysics experiment, we adjusted brightness of the flash in smaller step sizes (3.3 cd/m^2) and only after 4 consecutive 'not seen' responses or after 3 consecutive 'seen' (including all three "partially visible" levels: clear, almost clear & weak glimpse) responses.

For EEG experiment, we kept flash luminance constant throughout the four days based on the initial psychophysics experiment to ensure that observed differences in brain activity are related to stimuli perception (i.e. flash visibility) rather than their physical properties. Each participant completed 76 experimental runs over 4 testing days, resulting in 4104 trials (i.e. 608 trials for each flash present condition and 152 trials for each flash absent condition).

Inclusion criteria for psychophysics and EEG

For the psychophysics experiment, we limited the analysis to 72 subjects based on two exclusion criteria: first, we included only subjects that provided reliable visibility judgments as indicated by 'better than chance localization accuracy' (based on binomial test) for visible flashes (i.e. exclusion of 22 subjects). Second, to ensure reliable parameter estimation, we included only those participants who had at least 10 trials in each of the 2 (flash location: left vs. right) x 3 (sound location: left, middle, right) conditions, for both visible and invisible categories respectively (i.e. exclusion of 9 subjects). These exclusion criteria ensured that the computation of the ventriloquist effect was based on at least 60 trials.

For the EEG experiment, we invited subjects back from the initial psychophysics experiments, if their ventriloquist effect for invisible trials was equal to or greater than 0.05, and their flash localization accuracy for invisible trials was not higher than 56%. 31 out of 72 subjects were invited to participate in the EEG experiment, but only 18 agreed to complete all four EEG sessions. We included all participants that completed all four EEG sessions in the analysis.

EEG data acquisition and preprocessing

Continuous EEG signals were recorded from 60 channels using Ag/AgCl active electrodes arranged in 10-20 layout (ActiCap, Brain Products GmbH, Gilching, Germany) at a sampling rate of 1000Hz (for technical reasons, 6 sessions spread across 3 subjects were recorded at a sampling rate of 500Hz), referenced at FCz. Four electrodes were used for EOG recording (2 placed above and below right eye and two on the side of each eye). Impedances were kept below 10kΩ for EEG channels and 20kΩ for EOG channels. On each testing day, 3D coordinates of EEG electrodes were digitized using Polhemus Fastrack (Polhemus Corp., Colchester, US).

Preprocessing was performed using MATLAB R2016b (Mathworks, Natick, Massachusetts) and Fieldtrip toolbox (Oostenveld, Fries, Maris, & Schoffelen, 2011). Muscle artefacts and noisy channels (0.8 channels on average) were identified based on visual inspection and rejected. Continuous EEG signals were high-pass filtered to 0.1 Hz and low-pass filtered to 30Hz. Independent component analysis (ICA) was applied to correct for eye movements and heartbeat artefacts. Eye blink and heartbeat-related components were identified based on visual inspection of component topographies and time-courses. Between 2 and 6 ICA components were removed (3.1 components on average). Noisy channels were interpolated using weighted average of neighbouring channels, based on sensor positions from Polhemus recordings. Data were segmented into -0.15:0.65s epochs relative to target stimulus (i.e. flash-white noise burst stimulus) onset and re-referenced to average reference. After re-referencing, the FCz electrode was appended, so that signals from 61 channels were entered into the analysis. Trials containing artefacts were rejected. Furthermore, trials were rejected if they included an eye blink overlapping with presentation of the flash. Data was baseline corrected and downsampled to 60Hz.

Behavioural analysis for psychophysics and EEG experiment

The behavioural data of the initial psychophysics study and the EEG experiment were analysed and reported separately in the text. All figures show only the results from the EEG experiment.

For data analysis, we reduced the four visibility levels to two visibility levels: 1. Visible = 'clear image', 'almost clear image', 'weak glimpse' and 2. Invisible = 'not seen'. Further, we pooled over flashes in upper and lower fields given their negligible elevation (i.e. ± 1.2° visual angle).

For each participant, we coded their sound location responses as -1 for left, 0 for centre and 1 for right across trials. We estimated the *perceived sound location* for each of the 2 (flash location: left, right) x 3 (sound location: left, centre, right) conditions by averaging the localization responses across trials. Next, we averaged the perceived sound locations separately for trials where the flash was presented on the left and right and computed the difference in average perceived sound location for 'visual right' minus 'visual left' trials as the index of the spatial ventriloquist effect. A positive value of this index indicates that subject's perceived sound location shifted towards the visual stimulus location (i.e. attraction) and negative value indicates that it is shifted away from the visual stimulus location (i.e. repulsion). A ventriloquist effect of zero means that participants were not influenced consistently across trials by the location of the flash.

This difference in perceived sound location, i.e. the ventriloquist effect, was then used as the dependent variable for all subsequent analyses. If the visual signal location attracts the perceived sound location, we would expect the difference to be significantly greater than zero. Given our a priori directed hypothesis, all p-values are reported for right-tailed one sample t-tests.

The data analysis was performed in MATLAB (Mathworks, Natick, Massachusetts), Bayes Factors were computed in JASP (Marsman & Wagenmakers, 2017).

Spatial ventriloquism for visible and invisible trials

We investigated whether the ventriloquist effect was present independently for both invisible and visible flashes. Hence, we computed the ventriloquist effect separately for trials where visual signals were judged visible or invisible and tested whether the ventriloquist effect was significantly greater than zero in right-tailed one sample t-tests independently for visible and invisible trials.

Moreover, for the behavioural data from the four-day EEG experiment alone we also tested whether the ventriloquist effect still can be observed in invisible trials, when participants were not able to locate the flash. This analysis was not possible for the psychophysics experiment where data were limited to one single acquisition session.

Correlation between visible and invisible ventriloquist effects

We investigated whether participants that show a strong (resp. weak) ventriloquist effect for invisible trials also exhibit a strong (resp. weak) ventriloquist effect for visible trials by testing for a Pearson correlation between the ventriloquist effects for visible and invisible trials. A significant correlation provides initial evidence that the neural mechanisms and circuitries underlying the ventriloquist effects in the presence and absence of awareness may be at least partly overlapping.

Influence of question order on flash localization accuracy and ventriloquism

As described earlier we counterbalanced the task order (either: sound location – visibility – flash location or: flash location – visibility – sound location) across participants to account for the fact that the order of the questions can influence observer's performance accuracy on the different tasks. For instance, as a result of memory noise, the order of questions may influence observers' flash localization accuracy and the size of the ventriloquist effect.

As a consequence, the question order could also influence the inclusion of participants. For instance, we may include a smaller number of subjects that were presented with the sound task first, because their flash localization accuracy for visible trials may have been below our inclusion criterion. Conversely, we include less participants with flash task first, because their localization accuracy on the invisible trials may have been too high. To assess the effect of task order on the exclusion of participants we compared the number of excluded participants across the groups where the flash localization task was presented first vs. last in a Chi-square tests.

In the EEG study, we also compared flash localization accuracy and the size of spatial ventriloquism between the groups of observers that performed the flash localization as first and last task using a two-sample t-test.

EEG analysis

Using multivariate pattern analysis, we investigated: 1. how the location of the flash is encoded in neural activity (i.e. classification accuracy) and 2. how the location of the flash influences the neural encoding of the sound location (i.e. neural ventriloquist effect). Critically, we examined how these processes depend on whether observers i. rated the flash as visible or invisible and ii. located the flash correctly or incorrectly.

All multivariate analyses were performed in MATLAB (Mathworks, Natick, Massachusetts) using the CoSMoMVPA toolbox (Oosterhof et al., 2016) and Libsvm package (Chang & Lin, 2011). We trained a support vector machine (C=1) on single trial EEG activity patterns pertaining to i. the entire time window from 0 to 650ms poststimulus or ii. 50ms sliding time windows (i.e. 61 channels x 3 time samples, given a 60Hz sampling rate). EEG signals were z-normalized in each channel with normalization parameters from training set being applied to testing set.

Using one sided t-tests, we investigated whether flash decoding accuracy was better than chance or the neural ventriloquist was greater than zero based on the entire time window. For the sliding time window analysis, we report p-values corrected for multiple comparisons using the Threshold Free Cluster Enhancement procedure (Smith & Nichols, 2009) with sign-permutation test (based on 10000 iterations) as implemented in CoSMoMVPA (Oosterhof et al., 2016). Unless stated otherwise, results are reported for $p < 0.05$.

Flash localization

Using support vector classification, we investigated how the brain encodes flash location depending on observers' visibility and localization accuracy. We balanced the number of trials from each of the audio-visual conditions (4 flash locations x 3 sound locations) in each of the training and testing folds. To minimize confounding neural activity from sound processing we limited the analysis to 10 parieto-occipital EEG channels.

First, using a 10-fold cross-validation, we trained a linear SVM to classify left vs. right flash locations (pooling over up vs. down locations) separately for flashes judged visible and invisible.

Second, we trained a SVM to classify left vs. right flash location on visible trials and then generalized this SVM to invisible trials i. irrespective of flash localization accuracy and ii. separately for trials where observers located the flash correctly and incorrectly.

For the sliding time window analysis, we corrected for multiple comparisons in the entire time window for visible trials. For invisible trials, we only tested in the time window with significantly better than chance decoding accuracy for visible trials.

Sound localization and neural ventriloquism

To investigate how the location of a synchronous flash influences the neural encoding of the sound, we trained a support vector regression (SVR) model on unisensory auditory trials (left, middle, right) to learn the mapping from EEG activity patterns to sound location. We used this SVR model to predict the sound location from EEG activity patterns of audiovisual trials. We computed a neural ventriloquist effect by subtracting the decoded sound location for flash right trials from flash left trials (i.e. similar to our behavioural analysis) and assessed how this 'neural ventriloquist effect' depends on i. flash visibility, ii. flash localization accuracy (for invisible trials only, due to insufficient number of incorrect visible trials) and iii. the occurrence of a behavioural ventriloquist effect. Because these three factors can be assessed in a balanced fashion only for trials where the sound was presented in the middle, our neural ventriloquist analysis is limited to the sound centre trials only.

The neural ventriloquist effect was entered into two repeated measures ANOVAs (n.b. we acknowledge that we transformed dependent variables into independent factors): 2(visibility: visible vs. invisible) x 2(behavioural ventriloquist effect: present vs. absent) and second for invisible trials only 2 (flash localization accuracy: correct vs. incorrect) x 2(behavioural ventriloquist effect: present vs. absent).

ANOVAs and Bayesian statistics were computed in JASP (Marsman & Wagenmakers, 2017).

Influence of question order on decoding accuracy of flash location and neural ventriloquism

Similar to our behavioural analysis, we compared decoding accuracy of flash location and the neural ventriloquist between groups of subjects who performed flash or sound localization tasks first.

Results

Behavioural results

Spatial ventriloquism for visible and invisible trials

As expected, we replicated results of the previous experiments of (Delong et al., 2018) and observed a significant ventriloquist effect for visible and invisible conditions (see Table 4.1). Ventriloquism was significantly larger for visible comparing to invisible flashes as indicated by paired t-test (psychophysics: t(71) = 10.889, p < 0.001; EEG experiment: t(17) = 5.748, p < 0.001).

Importantly, we observed significant ventriloquist illusion, even for invisible trials, when subjects were not able to locate the flash.

Experiment	Condition	Mean VE ± SEM	t value	p value	Cohen's d ± 95% CI
Psychophysics, df = 71	Visible	0.53 ± 0.05	11.673	< 0.001	1.38 ± 0.36
	Invisible	0.07 ± 0.01	5.944	< 0.001	0.7 ± 0.34
EEG, df = 17	Visible	0.88 ± 0.15	5.880	< 0.001	1.39 ± 0.73
	Invisible	0.09 ± 0.02	5.160	< 0.001	1.22 ± 0.71
	Invisible Correct	0.13 ± 0.03	4.496	< 0.001	1.06 ± 0.7
	Invisible Incorrect	0.05 ± 0.01	4.652	< 0.001	1.1 ± 0.7

Table 4.1 Behavioural ventriloquist effect.

Correlation between ventriloquism for visible and invisible trials

We observed significant correlation between size of subjects' VE for trials judged as visible and invisible (psychophysics: Pearson's R = 0.404, p < 0.001, EEG experiment: Pearson's R = 0.73, p < 0.001).

Influence of question order on flash localization accuracy and ventriloquism

There was no effect of question order either on meeting inclusion criteria for analysis (i.e. having min. 10 trials per condition and better than chance flash localization accuracy for visible trials) (chi^2 = 0.023, p = 0.88) or on meeting inclusion criteria for EEG study (i.e. VE $_{\text{Invisible}}$ of at least 0.05 and accuracy for invisible trials smaller than 56%) (chi^2 = 0.510, p = 0.475).

Two sample t-tests did not indicate significant influence of question order on flash localization accuracy for either visible or invisible trials (p_{vis} = 0.905, t(16) = 0.121, BF_{01} = 2.407, p_{inv} = 0.956, t(16) = -0.056, BF_{01} = 2.417) or ventriloquist effect (p_{vis} = 0.117, t(16) = 1.659, BF_{01} = 0.978, p_{inv}

= 0.813, t(16) = 0.241, BF_{01} = 2.371, $p_{inv\ correct}$ = 0.986, t(16) = 0.018, BF_{01} = 2.419, $p_{inv\ incorrect}$ = 0.649, t(16) = 0.464, BF_{01} = 2.246).

EEG analysis

Flash localization

Full time window (0 – 650ms) classification yielded classifier prediction accuracy for left vs right flash locations as higher than chance both for visible and invisible conditions (see Table 4.2 for detailed results). Importantly spatial information generalized from visible to invisible trials. Accuracy of classifier trained on visible and tested on invisible trials was not only significantly better than chance, but even increased comparing to training on invisible trials. This improvement was not statistically significant (paired t-test: t(17) = -1.214, p = 0.241) , but we can definitely say that the classification algorithm trained on visible trials is at least as good or better than one trained on invisible (Bayesian paired, one sided t-test for Acc trained on visible < Acc trained on invisible BF_{01} = 8.151).

The algorithm trained on visible trials successfully predicted flash locations for invisible conditions both when participants were and were not able to locate the flash. We did not observe a significant difference in classifier performance between these conditions (paired t-test: t(17) = 0.655, p = 0.522) with the null hypothesis confirmed using Bayesian statistics (Bayesian paired t-test: BF_{01} = 3.401).

For visible trials, flash location decoding is significant starting from about 130ms poststimulus until the end of used time window (650ms). For trials judged invisible, significant classifier performance started 70ms later (see Figure 4.2B). However, when trained on visible trials, better than chance decoding starts at 150ms, indicating that relevant visual information in invisible condition is already present at that time. For both invisible correct and incorrect trials, we observed peak of decoding accuracy between 150 and 300ms, but after 300ms visual-spatial information was considerably diminished for incorrectly located flashes (as indicated by significant difference between conditions after that time).

Tested / Trained	**Mean Accuracy ± SEM %**	**t(17)**	**p value**
Visible / Visible	66.6 ± 2.3%	7	< 0.001
Invisible / Invisible	53.1 ± 0.8%	3.887	< 0.001
Invisible / Visible	54.3 ± 1%	4.157	< 0.001
Invisible Correct / Visible	54.9 ± 1%	4.823	< 0.001
Invisible Incorrect / Visible	54.2 ± 1.1%	3.846	< 0.001

Table 4.2 Classification accuracy for flash localization.

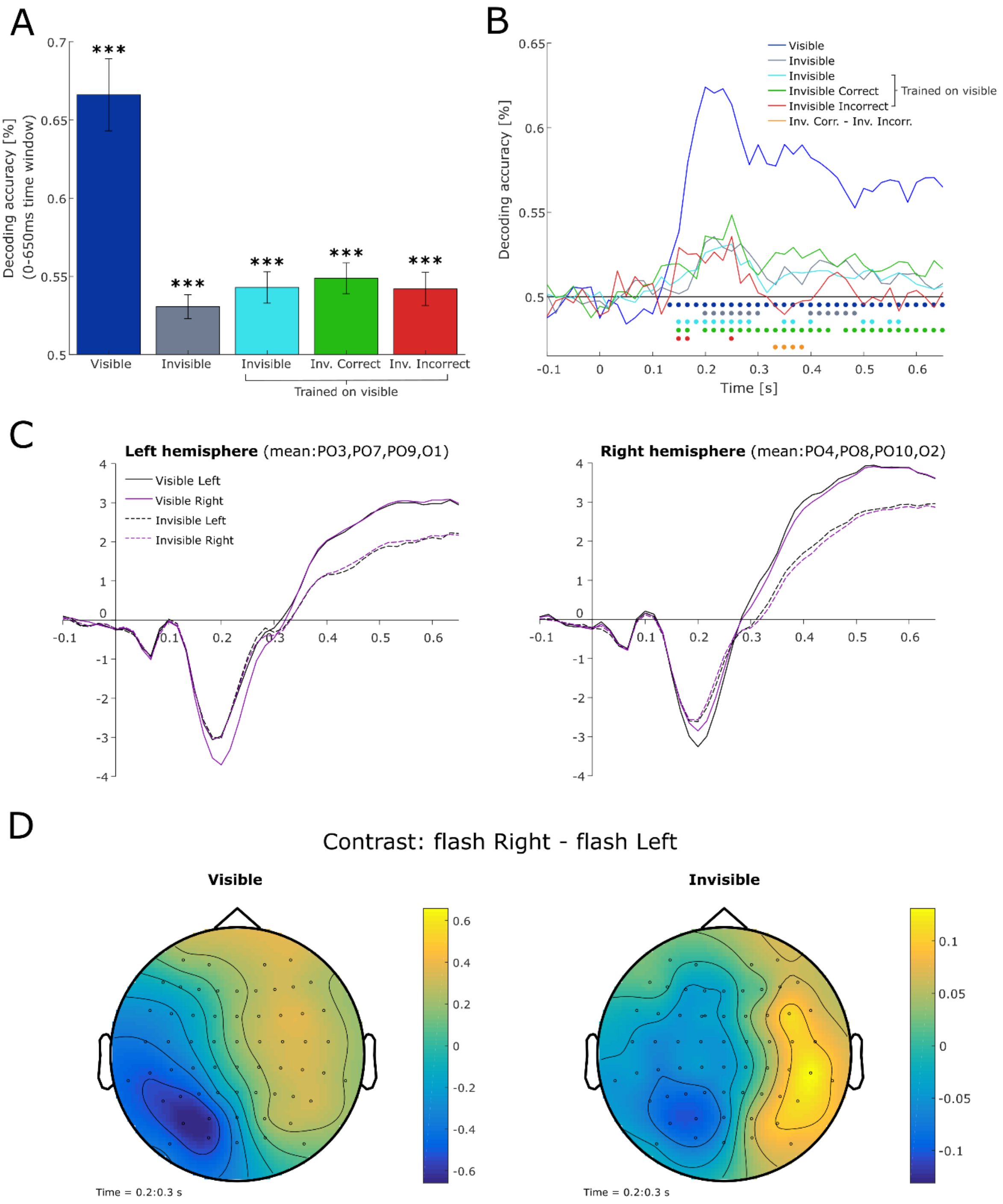

Figure 4.2 Visual-spatial representations. A. Bar plots showing classifier accuracy for Left vs Right flash location decoding (across subjects mean ± SEM). B. Timecourse of classifier's accuracy. Dots indicate classifier's performance significantly better than chance. C. ERPs for visible/invisible flashes presented in the right or left locations. Grand averages were computed by averaging all trials for each condition first within each participant, then across participants. D. Topographies of difference between flash Right and flash Left conditions, for visible and invisible trials. Shown for 200 - 300ms time window (peak of decoding accuracy).

In all figures: ***p < 0.001, **p < 0.01, *p < 0.05, n.s. p > 0.05.

Sound localization and neural ventriloquist illusion

Average sound locations predicted by support vector regression trained on unimodal sounds are shown in Figure 4.3B. Ventriloquist effect computed using these predictions was ~5 times smaller than behavioural VE, but nevertheless significant for both visible (see Table 4.3 for detailed results) and invisible trials. For incorrectly located invisible flashes we did not observe significant neural VE.

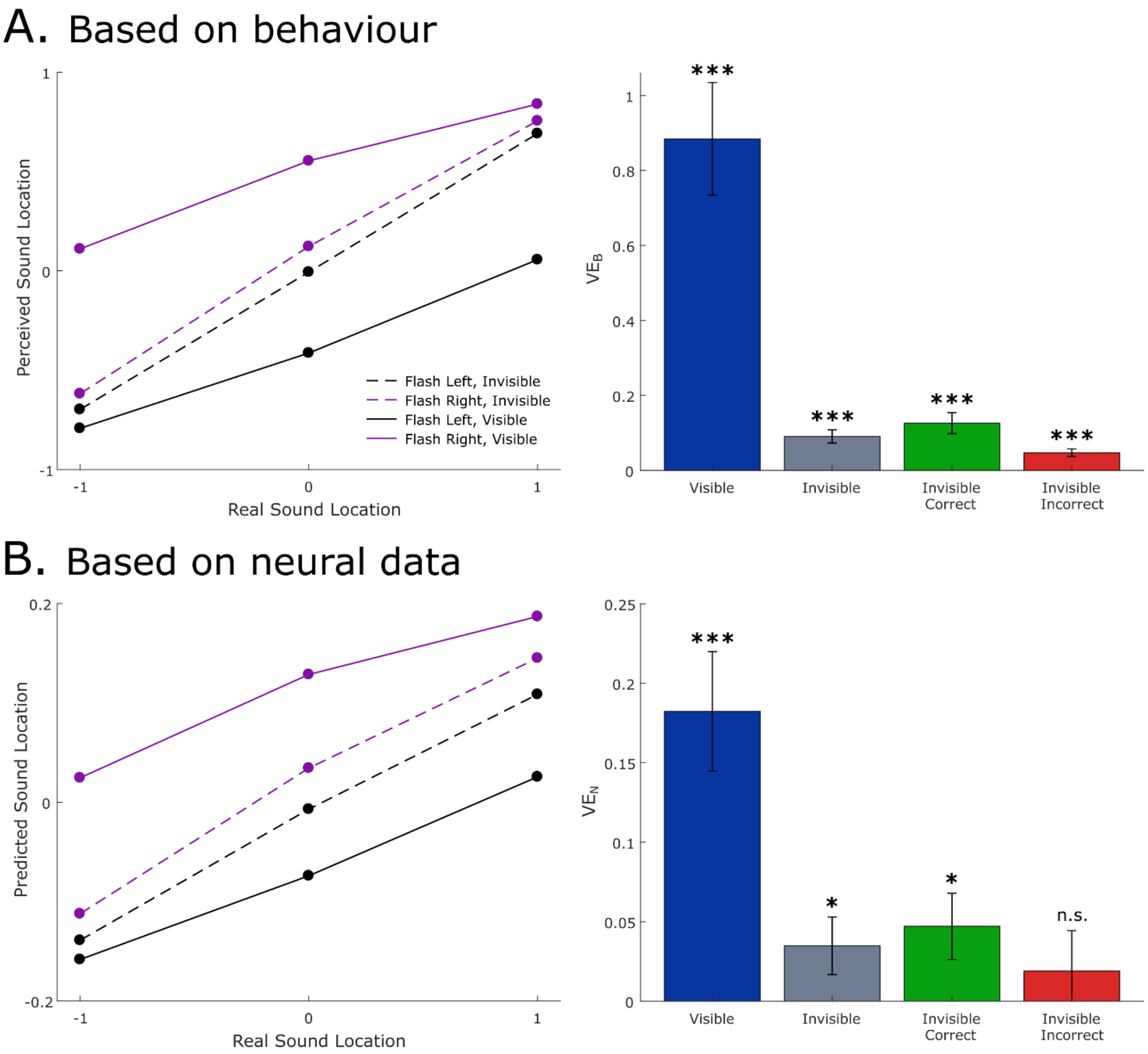

Figure 4.3 Ventriloquist illusion and sound perception. A. Mean reported sound locations for different AV conditions and ventriloquist effect computed for behavioural data ($VE_{B,}$ across subjects mean ± SEM). B. Mean sound locations predicted by classifier and ventriloquist effect computed using these predictions (VE_N, across subjects mean ± SEM). Please note different scales for behavioural and neural results.

Similar to behavioural results the neural ventriloquist effect is attenuated for invisible comparing to visible trials (paired t-test t(17) = 4.888, p < 0.001). Based on behavioural data however, it is not possible to determine whether this is a result of ventriloquist illusion occurring more often for visible trials or invisible flashes causing smaller shifts in sound perception. To address that we compared this neural ventriloquism for trials where participants did or did not experience the ventriloquist illusion.

Condition	Mean VE_N ± SEM	t(17)	p value	Cohen's d ± 95% CI
Visible	0.18 ± 0.04	4.848	< 0.001	1.14 ± 0.7
Invisible	0.03 ± 0.02	1.926	0.036	0.45 ± 0.66
Invisible Correct	0.05 ± 0.02	2.259	0.019	0.53 ± 0.66
Invisible Incorrect	0.02 ± 0.03	0.752	0.231	0.18 ± 0.65

Table 4.3 Neural ventriloquist effect.

For the sound centre condition, the neural VE was highly significant for trials with, but not without the ventriloquist illusion (see Table 4.4 for detailed results), for both visibility levels.

Condition	VE_B	Mean VE_N ± SEM	t(17)	p value	Cohen's d ± 95% CI
Visible	yes	0.23 ± 0.04	5.547	< 0.001	1.31 ± 0.72
	no	0.09 ± 0.07	1.315	0.103	0.31 ± 0.66
Invisible	yes	0.19 ± 0.06	3.019	0.004	0.71 ± 0.67
	no	-0.01 ± 0.03	-0.269	0.605	-0.06 ± 0.65
Invisible Correct	yes	0.23 ± 0.08	3.055	0.004	0.72 ± 0.67
	no	0.01 ± 0.04	0.236	0.408	0.06 ± 0.65
Invisible Incorrect	yes	0.15 ± 0.08	2.003	0.031	0.47 ± 0.66
	no	-0.02 ± 0.05	-0.442	0.668	-0.1 ± 0.65

Table 4.4 Neural ventriloquist effect for sound centre condition.

A repeated measures ANOVA has shown a significant effect of behavioural ventriloquism on the neural VE (F(1,17) = 13.481, p = 0.002), but not of visibility (F(1,17) = 2.58, p = 0.127). Moreover, Bayesian paired t-test confirmed that there was no difference in neural ventriloquism between visible and invisible trials when subjects reported ventriloquist illusion (BF_{01} = 3.275).

For invisible correct and incorrect conditions, neural ventriloquism is significantly greater than zero for illusion trials (see Table 4.4) and not for non-illusion trials. A repeated measures ANOVA has only shown an effect of behavioural ventriloquism (F(1,17) = 10.958, p = 0.004), but not of correct flash localization (F(1,17) = 1.446, p = 0.246). Bayesian paired t-test suggests that there was no difference in neural ventriloquism between correct and incorrect invisible trials when subjects reported ventriloquist illusion, but did not provide strong evidence (BF_{01} = 2.444).

The temporal dynamic of ventriloquism computed based on classifier predictions over time is shown in Figures 4.4 B&D. For visible and invisible trials with reported VE_B, the highest peak was observed between 250 and 450ms poststimulus. For visible trials without reported VE_B shift in predicted sound location (VE_N) becomes significant later – after 500ms poststimulus.

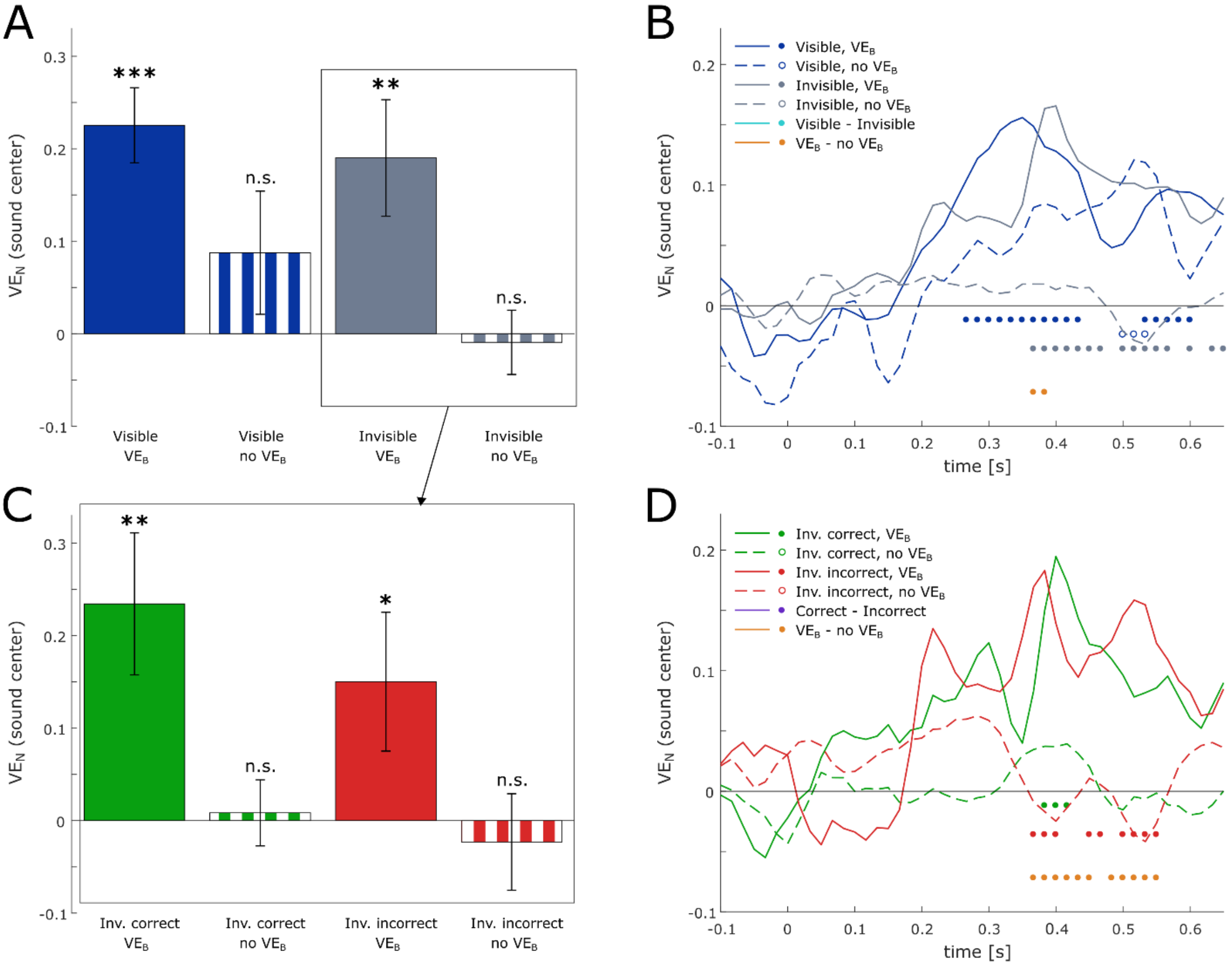

Figure 4.4 Neural ventriloquist illusion for sound centre location, for trials with or without reported ventriloquist illusion (VE$_B$). A&C Bar plots showing VE$_N$ (mean ± SEM) for classifier predictions based on the entire time window:0-650ms. B&D Time course of neural ventriloquism (VE$_N$). Dots indicate VE$_N$ significantly greater than zero. Temporal smoothing (average over 3 neighbouring data points – one on each side) was applied to plots B&D for visualization purposes (statistical analyses were performed for non-smoothed data). A&B For visible and invisible flashes. C&D For correctly/incorrectly located invisible flashes.

Influence of question order on decoding accuracy of flash location and neural ventriloquism

We did not observe a significant effect of question order on flash location decoding (visible: p = 0.654, t(16) = 0.457, BF$_{01}$ = 2.251, invisible: p = 0.913, t(16) = 0.111 , BF$_{01}$ = 2.409, invisible trained visible: p = 0.416, t(16) = - 0.834, BF$_{01}$ = 1.905, invisible correct trained visible: p = 0.984, t(16) = 0.021, BF$_{01}$ = 2.419, invisible incorrect trained visible: p = 0.953, t(16) = 0.06, BF$_{01}$ = 2.417) or neural ventriloquist illusion (visible: p = 0.702, t(16) = 0.389, BF$_{01}$ = 2.296, invisible: $p = 0.327$, $t(16) = 1.01$, $BF_{01} = 1.709$, invisible correct: $p = 0.173$, $t(16) = 1.427$, $BF_{01} = 1.227$, $p = 0.718$, $t(16) = 0.367$, $BF_{01} = 2.309$).

Discussion

This psychophysics-EEG study used dynamic continuous flash suppression (Maruya et al., 2008; Tsuchiya & Koch, 2005) in a spatial ventriloquist paradigm to investigate the neural mechanisms by which invisible flashes can influence where we perceive sounds. Importantly dynamic flash suppression obliterated visual awareness only on a fraction of trials. This allowed us to compare the impact of physically identical flashes that were i. visible or invisible and ii. located correctly or incorrectly on sound perception.

Our behavioural results show that observers perceived the sound shifted (i.e. biased) towards the true flash location. While this spatial ventriloquism was stronger for visible flashes, a robust ventriloquist effect was also present for invisible flashes. Critically, a behavioural ventriloquist effect was also preserved for flashes that were mislocated by observers. In other words, even when observers did not have reliable spatial information for accurate spatial discrimination along elevation, the perceived sound location was biased towards the true flash location along the azimuth (Figure 4.3). Our results are consistent with a recent psychophysics study showing a ventriloquist illusion for invisible flashes that could not be located better than chance (Delong et al., 2018). By contrast, the McGurk illusion has previously been shown to falter when visual signals were rendered invisible (Ching et al., 2019; Palmer & Ramsey, 2012). These differences in susceptibility to 'awareness' manipulations may be explained by the fact that the ventriloquist illusion relies on low level spatiotemporal binding, while the McGurk illusion requires integration of complex audiovisual features such as visemes (i.e. articulatory facial movements) and phonemes. Indeed, mounting research has shown that higher order processing is obliterated when stimuli are rendered unconscious making the integration of complex audiovisual features less likely (Biderman & Mudrik, 2018; Faivre, Dubois, Schwartz, & Mudrik, 2019; Heyman & Moors, 2014; Moors, Boelens, van Overwalle, & Wagemans, 2016). Interestingly though, a recent study of Scott et al. 2019 suggests that unconscious learning of semantic associations between auditory and visual stimuli is possible. While associative learning for unconscious bimodal pairs, shown in their experiment, pose a challenge for GWT, it does not imply multisensory integration and could be driven by different mechanisms.

At the neural level, we first assessed how the neural encoding of flash location differs depending on observers' visibility ratings and flash localization accuracy. As shown in Figure 4.2, the left/right flash location was decoded successfully from EEG activity patterns between 150-300ms for both visible and invisible flashes but with a substantially greater decoding accuracy for visible than invisible flashes. This is not surprising, because invisible flashes elicited attenuated evoked responses and topographies comparing to visible flashes (Figure 4.2 C & D). Critically, after 300ms the flash location along the azimuth could be decoded significantly better than chance for invisible flashes only for trials when observers were able to locate the flash accurately.

This 'chance spatial decoding' after 300ms for erroneous localization responses may be surprising in light of recent research showing that stimulus location can be decoded better than chance from 270-800ms even on erroneous trials in an 8 alternative forced choice spatial task (Salti et al., 2015). Likewise, King et al. were able to decode the orientation of a Gabor patch on invisible trials until 1400ms poststimulus – though this study did not dissociate between trials with correct and incorrect responses (King, Pescetelli, & Dehaene, 2016). Critically, the decoded information in both studies was relevant for observers' visual decisions. By contrast, in our study we decoded flash location along the azimuth rather than elevation, which was the task-relevant dimension for observers' visual spatial discrimination task. Indeed, (King et al., 2016) also showed that task-irrelevant information such as a Gabor patch's spatial frequency could not be decoded successfully from MEG activity pattern after 230ms poststimulus. Thus, unconscious information that is relevant for task-performance may be encoded in more temporally stable activity patterns than task-irrelevant information potentially via top-down attentional mechanisms. Critically, both studies employed backward masking as visual suppression method whereas Continuous Flash Suppression was used in our case. CFS has been shown to be a stronger suppression method, and to disrupt visual processing at earlier stage (Izatt et al., 2014; Peremen & Lamy, 2014; Yuval-Greenberg & Heeger, 2013).

In summary, the results discussed so far suggest that visual spatial information along the azimuth can be decoded successfully from EEG activity between 150-300ms for invisible flashes irrespective of observers' flash localization accuracy. After 300ms, however, spatial azimuthal information rapidly evaporates for 'invisible' flashes that are mislocated by observers (Figure 4.2B). This suggests that the behavioural ventriloquist illusion for erroneously localized flashes is mediated by audiovisual interactions that occur early at about 150-300ms before information about flash location dissipates in the visual system.

To further characterize how visual signals impact sound processing we computed the neural ventriloquist effect, i.e. the shift of the decoded sound location towards the spatially incongruent flash for the entire time window from 0 to 650ms depending on i. flash visibility, ii. flash localization accuracy and iii. the occurrence of a perceptual ventriloquist illusion. As shown in Figure 4.4A, visible flashes always elicited at least a small shift in sound perception– however this neural ventriloquist effect was only significant, when observers experienced the ventriloquist illusion. Together these results suggest that the neural ventriloquist effect is closely linked to the emergence of a perceptual ventriloquist illusion.

Critically, the neural ventriloquist effect evolved with similar time courses irrespective of flash visibility. For both visible and invisible flashes, the neural ventriloquist effect started at about 250ms and peaked at 350ms poststimulus (Figure 4.4), which converges with recent EEG research on supraliminal spatial ventriloquism. For instance, multivariate EEG decoding revealed a neural ventriloquist effect that culminated at about 350ms (Aller & Noppeney, 2019). Likewise, observers' ERPs differed at about ~270ms poststimulus depending on whether they experienced a perceptual ventriloquist illusion (Bonath et al., 2007). Interestingly, we also observed significant shift in neural VE for visible trials, when observers did not experience the illusion, although this shift occurred at later time – abut 500ms poststimulus. These results suggest that reliable (i.e. visible) spatial representations automatically impact neural processing of sounds at this later processing stage; yet, because of additional stochastic effects this neural ventriloquist effect does not necessarily manifest itself at the behavioural level (i.e. no perceptual ventriloquist illusion). By contrast, a significant neural ventriloquist effect was observed for invisible flashes only when observers experienced the ventriloquist illusion. This suggests that the neural ventriloquist effect is more closely related to the emergence of a perceptual ventriloquist illusion for invisible than visible flashes. Collectively, the timecourse of the flash decoding accuracy and neural ventriloquist effect suggest that invisible flashes influence sound processing via early multisensory interactions, ultimately leading to neural and perceptual ventriloquist effects that are indistinguishable from those observed for visible flashes (see Figure 4.5 showing proposed audio-visual integration model). Indeed, a final common pathway for ventriloquism in the absence and presence of subjective awareness is also supported by a correlation between the visible and invisible ventriloquist effects. Participants that frequently experienced the ventriloquist illusion for visible flashes also reported ventriloquist illusions more often for invisible flashes.

In conclusion, our findings unravel how a subliminal flash can change where we perceive a sound. We suggest that early multisensory interactions mediate the influence of unconscious signals in one sensory modality on neural and ultimately perceptual representations in another sensory modality. (see Figure 4.5, right, model 2). These results demonstrate that consciousness is not a generic prerequisite for information integration, not even in situations where the signals stem from different sensory channels – thereby providing new challenges for the Global Workspace Theories.

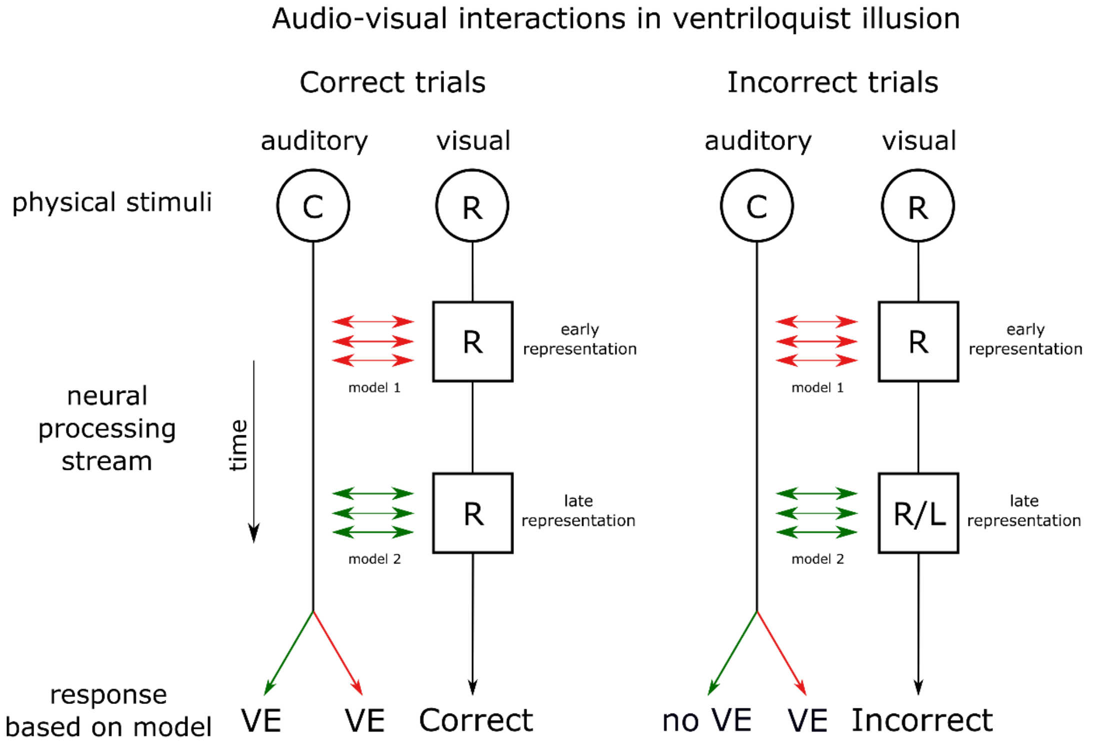

Figure 4.5 Proposed models of audio-visual interactions in ventriloquism. Models of how invisible flashes can influence sound perception: Left. On trials where the flash is correctly localized visual spatial information may be available in early and late processing stages to influence neural spatial representations of sound. Right. On trials where the flash is incorrectly localized visual spatial information may rapidly dissipate in the visual system, but can influence auditory spatial processing via early interactions, so that the neurally encoded and reported sound location is shifted towards the true flash location (i.e. neural and behavioural ventriloquist effect).
Spatial locations: L = left, C = centre, R = right. VE = ventriloquist effect

CHAPTER 5. THE ROLE OF SEMANTIC CONGRUENCY AND AWARENESS IN SPATIAL VENTRILOQUISM

Patrycja Delong, Uta Noppeney

Contributions:

PD and UN designed the study. PD acquired the data. PD analysed the data under supervision of UN. PD and UN wrote the abstract. PD wrote the manuscript.

Acknowledgements:

We thank Alyssa Sevilla and Uzoma Ezihe for help with data acquisition, and Nicholas Berry and Jasmine Hawkins for help with pilot data acquisition (not included here).

Results of this chapter were presented at:

International Multisensory Research Forum 2018, Toronto, Canada, 2018

Poster: “The role of semantic congruency and awareness in spatial ventriloquism”

Abstract

The extent to which signals from different senses can interact in the absence of awareness is controversial. Models of global workspace predict that unaware signals are confined to processing in low-level sensory areas and thereby are prevented from interacting with signals from other senses in higher order association areas. Numerous previous studies have shown that sounds can boost unaware visual stimuli into perceptual awareness depending on the semantic congruency between auditory and visual signal. Yet, it is unclear to what extent unaware visual stimuli can influence our perception of the aware sounds. Using spatial ventriloquism, the current study investigated whether the unaware visual stimuli can influence where we perceive sounds depending on the semantic congruency of the audiovisual signals.

In a spatial ventriloquist paradigm, participants were presented with object pictures in synchrony with source sounds. In a 2 x 2 factorial design we manipulated the spatial and semantic congruency of the audiovisual signals. Across two experiments, we presented the object pictures without and with forward-backward masking. Participants reported the sound location, semantic category (exp. 2 only) and picture visibility (exp. 2 only).

We observed a significant ventriloquist effect irrespective of visual masking. Yet, semantic congruency profoundly modulated the size of the spatial ventriloquist effect only for unmasked stimuli. Nevertheless, as previously reported semantically congruent sounds increased the visibility of masked pictures and increased semantic categorization accuracy.

Our results suggest that sounds boost masked pictures into participants' awareness depending on semantic congruency. Yet, invisible pictures do not bias perceived sound

location depending on semantic congruency. The latter finding suggests that semantic congruency does not modulate audio-visual integration in the absence of awareness.

Introduction

Integrating information from different sensory inputs is necessary to create a unified percept of multisensory environment. It is debatable to what extent crossmodal interactions can occur in the absence of awareness. Global Workspace Theory implies that such interactions should be only possible for consciously aware information, where processing of unaware stimuli should be limited to local sensory regions (Baars, 2002, 2005). Recent studies however, have provided initial evidence contradicting the GWT predictions and showing that an unconscious flash can influence processing of a conscious sound (Delong et al., 2018) and crossmodal associative learning is possible between two unconscious signals (Scott et al., 2018).

Interestingly, previous studies showed ventriloquist (Delong et al., 2018), but not McGurk effects in the absence of visual awareness (Ching et al., 2019; Palmer & Ramsey, 2012). In the ventriloquist illusion, spatial sound perception is shifted towards a simultaneously presented visual stimulus (Thurlow & Jack, 1973), where in McGurk illusion, perceived auditory syllable is altered by a video showing incongruent lip movements (McGurk & MacDonald, 1976). Critically, both phenomena indicate multisensory integration of audio-visual stimuli. Why can only the ventriloquist effect, but not McGurk, be caused by visual stimuli obliterated from awareness? This discrepancy could be due to the type of information the illusion relies on, with the ventriloquist effect dependent on low-level spatial cues and McGurk on integration of complex audiovisual features (phonemes and visemes). Multiple experiments have shown that higher level processing was abolished for unconscious stimuli (Biderman & Mudrik, 2018;

Faivre et al., 2019; Heyman & Moors, 2014; Moors et al., 2016). Lack of high-level unconscious processing should also prevent crossmodal interactions dependent on semantic congruency of unconscious stimuli in Scott's et al. experiments (Scott et al., 2018). Notably, the audio-visual associative learning shown in their studies is not evidence for multisensory integration, as it does not involve creation of a unified percept and therefore could be driven by a different mechanism than the McGurk effect.

In the present study, we employed forward-backward masking and the ventriloquist illusion to investigate whether audio-visual integration can be affected by semantic congruency of visual stimuli, which are rendered invisible.

It has been shown that semantically congruent supraliminal sound can facilitate processing of subliminal visual stimuli depending on semantic correspondences (Y.-C. Chen & Spence, 2010; Y.-C. Chen et al., 2011; Cox & Hong, 2015; Hsiao, Chen, Spence, & Yeh, 2012; M. Lee et al., 2015). Please note, that influencing subliminal by supraliminal signals is permitted by GWT: aware signals can travel through the global workspace and affect local processing of the unaware signals. It is also well established that semantic congruency between aware stimuli can modulate multisensory integration (for review see: Tsilionis & Vatakis, 2016). Interestingly though, a few studies have reported that they do not influence ventriloquist illusion. While early work of Jackson (Jackson, 1953) suggested the importance of realism in ventriloquism (these results have to be interpreted with caution as realistic and non-realistic conditions were examined in separate experiments, using different auditory stimuli and different angles of audio-visual discrepancies, which makes them incomparable), later studies showed no difference between VE for voice and upright/inverted faces (Bertelson et al., 1994; Colin et al.,

2001) and no difference in ventriloquism aftereffects between semantically congruent and incongruent audio-visual pairs (Radeau & Bertelson, 1977, 1978). At the same time, a modulatory effect of semantic congruency was shown in a recent study that employed a ventriloquist paradigm, in which talking faces were presented bilaterally (Kanaya & Yokosawa, 2011).

The lack of semantic modulation of VE for unilateral presentations could be a result of ceiling effects – if ventriloquist illusion for incongruent images is already large and additional binding cues (semantic congruency) cannot increase the size of the illusion. To investigate that possibility, in the current study we presented the visual stimulus either unilaterally or bilaterally - with distractor image presented in the opposite hemifield. Importantly, we first performed a psychophysics study without visual masking to confirm that semantic modulation of ventriloquism will occur in our paradigm for supraliminal images.

Methods

Participants

After giving informed consent, 44 healthy young adults (mean age ± std: 20.9 ± 5.7 years, range: 18-47 years, 6 male, 8 left-handed, 2 ambidextrous) took part in experiment 1, 44 subjects (mean age ± std: 20.9 ± 2.2 years, range: 18-30 years, 7 male, 4 left-handed, 2 ambidextrous) in experiment 2. Out of those, 12 subjects took part in both experiments. The study was performed in accordance with the principles outlined in the Declaration of Helsinki and was approved by the local ethics review board of the University of Birmingham. Sample size was determined assuming a medium effect size (Cohen's d = 0.5, for paired, two-tailed t-test), so it would be possible to detect even moderate influence of semantic modulation on the ventriloquist effect. For statistical power of 0.9, we obtained n = 44.

Stimuli and apparatus

Visual stimuli were a selection of six pictures (bird, car, dog, guitar, phone and daffodil) from Bank of Standard Stimuli database (Brodeur, Dionne-Dostie, Montreuil, & Lepage, 2010; Brodeur, Guérard, & Bouras, 2014), normalized for their familiarity. Images were displayed for 24ms on grey background of mean luminance 11 cd/m2. On each trial, a square image (5 visual degree side size) was centred at ±2.5 visual angle along the azimuth from the middle of the screen.

Auditory stimuli were five sounds (bird, car, dog, guitar, phone) downloaded from http:// www.findsounds.com (on 26/07/2017). The sounds were edited to start from the beginning of the sound file and to last for 150ms. Peak amplitudes of all the sounds were equalized with Audacity software (http://audacityteam.org). The sounds were presented via circumaural headphones (Sennheiser HD 280 Pro) and ranged in loudness from 66 to 75 dB SPL. To create a virtual auditory spatial signal, the sounds were convolved with spatially specific head-related transfer functions (HRTFs) thereby providing binaural (interaural time and amplitude differences) and monoaural spatial filtering signals. Recordings from MIT Media Lab database (Gardner & Martin, 1995) were used to interpolate HRTFs for desired spatial locations.

Psychophysical stimuli were generated and presented on a PC running Windows XP using the Psychtoolbox version 3.0.11 (Brainard, 1997; Kleiner, Brainard, & Pelli, 2007) running on MATLAB R2014a (Mathworks, Natick, Massachusetts).

Participants sat in a dimly lit room in front of a computer screen at viewing distance of 90cm. Visual stimuli were presented on a CRT monitor at a frame rate of 85 Hz (NVIDIA Quadro FX 380 graphics card). Auditory stimuli were digitized at a sampling rate of 44.8 kHz via a Sound Blaster Z SB1500 sound card.

Experimental design and procedure

In a spatial ventriloquist paradigm, participants were presented with an image in one of two locations: left or right (±2.5°) with semantically congruent or incongruent sound originating from left or right (±2.5°). In the congruent condition, image was presented together with corresponding auditory stimulus (5 congruent stimulus pairs). In the incongruent condition, image was presented with one of the four other auditory stimuli (20 combinations of incongruent pairs). The number of semantically congruent and incongruent presentations was equal.

Experiment 1

Audio visual stimuli were presented in 2 (spatial congruency: collocated, disparate) x 2 (semantic congruency: congruent, incongruent) x 2 (visual presentation mode: unilateral, bilateral) factorial design.

Each trial started with presentation of a fixation cross for 1000ms. Next, a target image was displayed for a duration of 24ms, followed by a white screen with fixation cross presented for 300ms. Sound was presented in synchrony with the picture. In bilateral presentation mode, distractor image (daffodil) was displayed in hemifield opposite to the target picture (see Figure 5.1A).

After each presentation participants were asked to locate the sound (left vs right), by shifting the cursor to the relevant answer on the screen (selected answer was highlighted) and pressing the left mouse button. The response screen was presented until an answer was provided or up to a maximum of 5 seconds.

The experiment consisted of two blocks: 320 trials with unilateral and 320 trials with bilateral image presentation. The order of the blocks was counterbalanced across subjects.

Experiment 2

Experiment 2 was identical to experiment 1, but we used backward-forward masking to suppress awareness of visual stimuli. Rectangles filled with coloured, dynamically moving Mondrians (similar as in (Delong et al., 2018; Maruya et al., 2008)) were used as a mask.

Prior to the experiment, subjects performed a practice session consisting of 5 unmasked visual trials, where they were asked to identify the image. If accuracy was lower than 100%, the practice session was repeated. This was to familiarise subjects with the combined task of reporting awareness and image identification.

Each trial started with presentation of the fixation cross for 800ms, followed by presentation of the mask for 200ms. Next, the target image was displayed on the screen for 24ms. The sound was presented in synchrony with the picture. Immediately after the visual stimulus, the mask was again presented for 300ms.

After each presentation, participants were asked to report: 1. the sound location and 2. the image visibility (using 4 level Perceptual Awareness Scale (Ramsøy & Overgaard, 2004)) along with its category. Visibility and picture category were reported at the same time – subjects selected the relevant visibility level and one of the semantic categories (see Figure 5.1B) using mouse cursor (selected box was highlighted) and confirmed by pressing the left mouse button. Response screens were presented until the answer was provided or up to a maximum of 5 seconds, starting from the moment that the sound location screen has been shown.

Participants completed 2x 640 trials in two blocks with unilateral or bilateral image presentation. The order of the blocks was counterbalanced across subjects.

Data analysis

Mean values for each participant and condition were computed in MATLAB R2016b (Mathworks, Natick, Massachusetts). Analysis of variance was performed in SPPS version 23 (IBM Corporation, Armonk NY). Significant interactions were further characterized by computing simple main effects for each factor. When analysis of simple effects did not explain the interaction (i.e. all simple main effects were significant and in the same direction) paired two-tailed t-tests were used to test for significant difference in size of those effects between conditions.

Experiment 1

To assess the impact of semantic congruency on the ventriloquist illusion, we employed a repeated measures ANOVA to compare sound localization accuracy between experimental condition. First, if participants experienced ventriloquism, they would perform worse in the spatially incongruent condition comparing to when auditory and visual stimuli were collocated, and we would observe a main effect of spatial congruency. Second, if semantic congruency can modulate the strength of the ventriloquist illusion, we would expect to see an interaction between spatial and semantic congruency. Third, if the semantic modulation is stronger for bilateral than unilateral presentations, we should observe an interaction between all 3 factors: spatial congruency, semantic congruency and presentation type. Finally, to directly compare the influence of semantic congruency on ventriloquism for unilateral and bilateral presentation (if an interaction between all 3 factors is observed), we first computed VE as a difference between sound localization accuracy for collocated and disparate stimuli and performed a paired t-test on the difference between semantically congruent and incongruent trials.

Experiment 2

Just as for experiment 1, we evaluated the influence of semantic congruency on the ventriloquist effect using a repeated measures ANOVA. Additionally, we investigated the impact of spatial and semantic congruency on image identification accuracy and mean visibility rating also employing repeated measures ANOVAs. Average visibility ratings were computed by assigning numerical values from 0 to 3 to four levels of Perceptual Awareness Scale (No Image = 0, Weak Glimpse = 1, Almost Clear = 2, Clear Image = 3).

Moreover, we performed the same analyses of sound localization and picture identification accuracy using visibility as an additional factor with two levels (visible or invisible). Trials with "No Image" response were considered invisible and the other three visibility ratings (Weak Glimpse, Almost Clear, Clear Image) were considered visible. These analyses were limited to participants, which had at least 10 trials that were judged as visible and at least 10 that were judged invisible in each of the 2 x 2 x 2 conditions (2 (spatial alignment: collocated, disparate) x 2 (semantic congruency: congruent, incongruent) x 2 (visual presentation mode: unilateral, bilateral)).

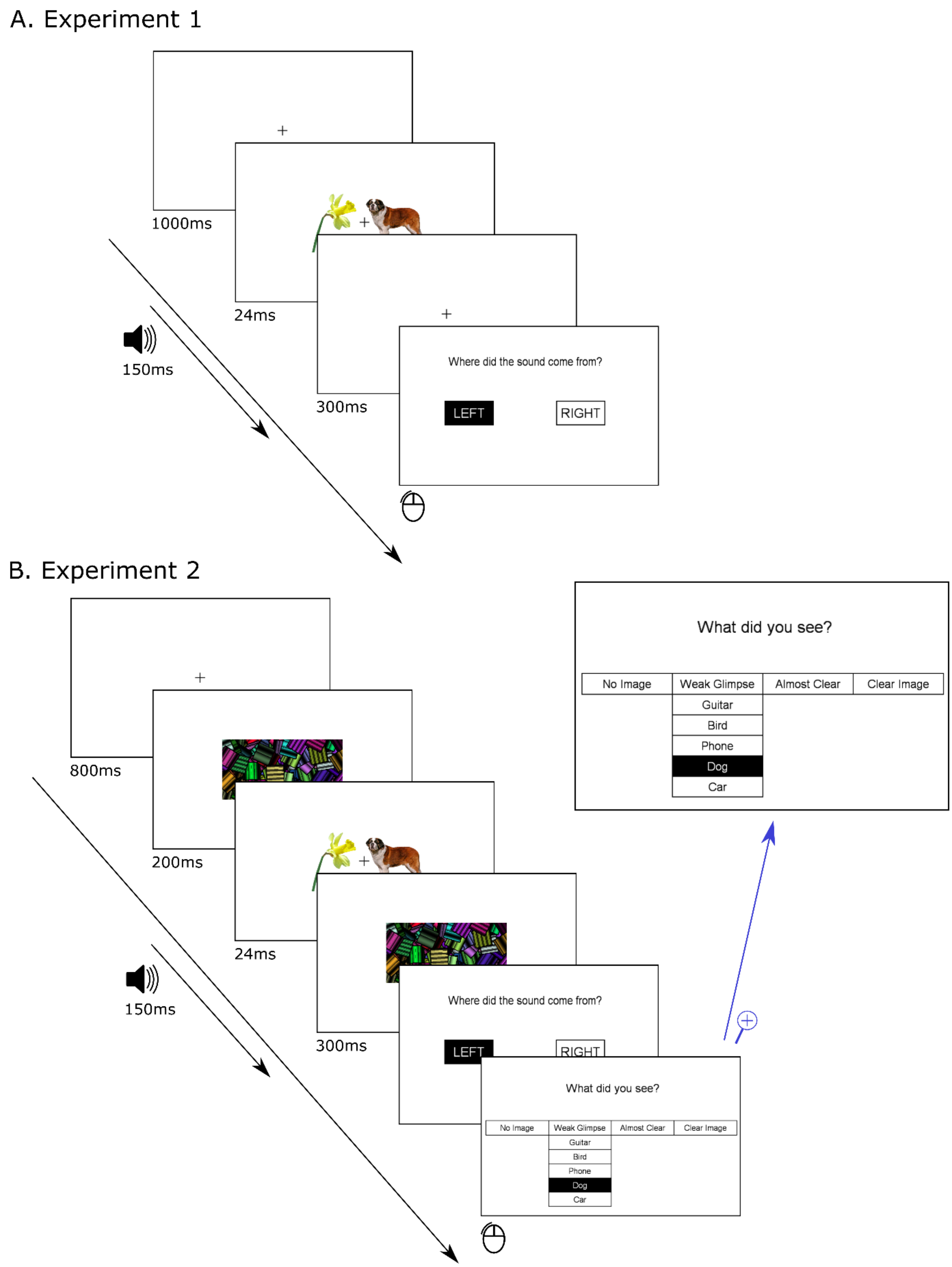

Figure 5.1 Example trials for bilateral presentations (i.e. with the distractor image presented on the opposite side to the visual target). In unilateral visual presentations, the picture of the daffodil was not shown. A. Experiment 1 – without visual masking. B. Experiment 2 – with forward-backward masking.

Results

Experiment 1

First, for the unmasked pictures we observed significant main effects for all tested factors influencing sound localization accuracy (F and p-values are listed in Table 5.1, mean accuracies ± SEM are shown in Figure 5.2A). Sound localization accuracy was higher for collocated comparing to disparate audio-visual pairs, which shows multisensory integration dependent on spatial congruency. This can be interpreted as a result of either multisensory enhancement (improved performance for audio-visual trials), or ventriloquist illusion (shift in perception towards disparate visual stimulus), or, which is most likely, a combination of both processes. Accuracy was lower for semantically congruent comparing to incongruent trials and for unimodal comparing to bimodal presentation. Second, we observed significant interactions between all tested factors. Sound localization was better for semantically congruent than incongruent pairs if they were collocated (simple main effect $p < 0.001$), but worse if they were spatially disparate (simple main effect $p < 0.001$). This interaction proved that ventriloquism can be modulated by semantic congruency. Spatial ventriloquism was also influenced by presentation mode (all simple main effects significant with p values < 0.001), with larger effects of spatial congruency (i.e. ventriloquist effect computed as difference between collocated and disparate conditions) for unilateral than bilateral visual presentation (paired t-test: $t(43) = 8.629$, $p < 0.001$). Finally, we observed an interaction between all three factors, where the simple effect of semantic congruency was not significant for collocated audio-visual pairs for unilateral presentation (simple main effect $p = 0.073$), with simple main effects significant in all the other conditions (all with $p < 0.001$). To directly compare the semantic modulation of ventriloquism in uni- and bilateral presentations, we took sound localization accuracy for disparate trials, subtracted it from accuracy in collocated trials, and then computed the difference between semantically congruent and incongruent conditions. This difference was significantly greater for bilateral presentations as indicated by a paired t-test (t(43) = -7.051, p < 0.001).

Factor	Sound Localization Accuracy – unmasked pictures (Exp. 1)			Sound Localization Accuracy – masked pictures (Exp. 2)		
	df	F	p	df	F	p
Picture number	1, 43	32.008	**< 0.001**	1, 43	3.936	0.054
Spatial congruency	1, 43	203.728	**< 0.001**	1, 43	21.41	**< 0.001**
Semantic congruency	1, 43	46.355	**< 0.001**	1, 43	0.454	0.504
Picture number * spatial congruency	1, 43	74.454	**< 0.001**	1, 43	0.788	0.38
Picture number * semantic congruency	1, 43	4.666	**0.036**	1, 43	0.5	0.483
Spatial congruency * semantic congruency	1, 43	58.458	**< 0.001**	1, 43	0.019	0.892
Picture number * spatial congruency * semantic congruency	1, 43	49.722	**< 0.001**	1, 43	1.338	0.254

Table 5.1 Results of repeated measures ANOVA: influence on sound localization accuracy.

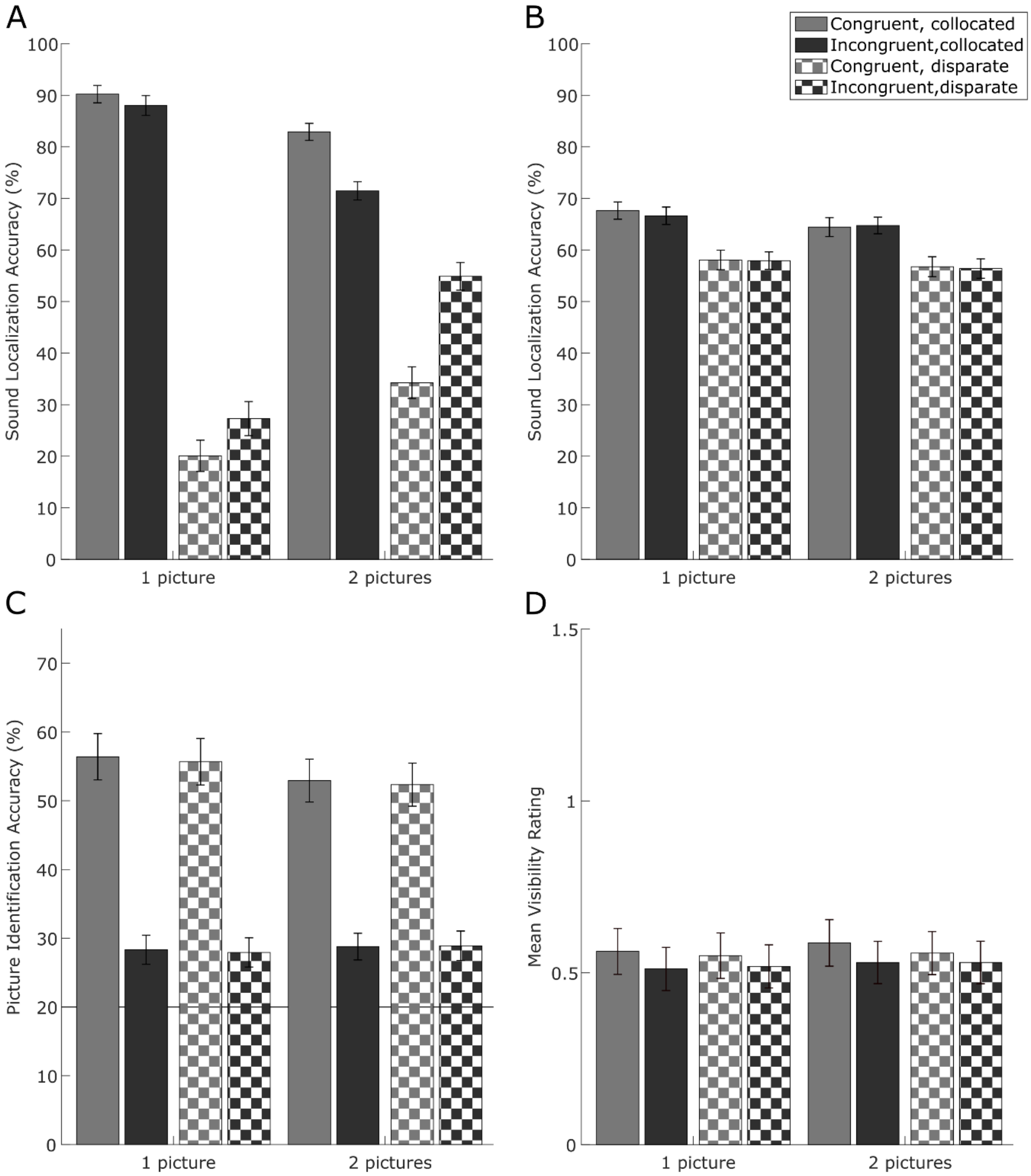

Figure 5.2 Sound localization, picture identification and visibility ratings dependent on visual presentation type (unilateral – 1 picture, bilateral – 2 pictures) and spatial alignment and semantic congruency of audio-visual stimuli. Bar plots showing across subjects mean ± SEM. A. Sound localization accuracy (Left vs Right) in Experiment 1 – without visual masking. B,C & D show results of experiment 2 – with backward-forward masking. B. Sound localization accuracy. C. Picture identification accuracy (choice of one of five pictures, horizontal line marks chance level of 20%). D. Mean visibility rating transformed to numerical values: 0 – Not Seen, 1 – Weak Glimpse, 2 – Almost Clear, 3 – Clear Image.

Experiment 2

Analysis for masked pictures

In the experiment using backward - forward masking, spatial congruency was the only factor that significantly modulated sound localization accuracy (F and p-values are listed in Table 5.1, mean accuracies ± SEM are shown in Figure 5.2B). Therefore, for masked pictures, we observed multisensory integration (ventriloquist effect), but it was no longer influenced by semantic correspondences or visual presentation type (unilateral vs bilateral). One could think that the ventriloquist illusion for bilateral presentation is necessarily dependent on semantic congruency, which in fact has a much simpler explanation. It has been shown that the more salient of the two synchronously presented, disparate visual attractors produces ventriloquism (Bertelson, Vroomen, et al., 2000).

Factor	Picture Identification Accuracy			Mean Visibility Rating		
	df	F	p	df	F	p
Picture number	1, 43	1.336	0.254	1, 43	0.34	0.563
Spatial congruency	1, 43	1.858	0.18	1, 43	2.738	0.105
Semantic congruency	1, 43	42.697	**< 0.001**	1, 43	14.305	**< 0.001**
Picture number * spatial congruency	1, 43	0.286	0.596	1, 43	2.427	0.127
Picture number * semantic congruency	1, 43	2.17	0.148	1, 43	0.008	0.927
Spatial congruency * semantic congruency	1, 43	0.53	0.47	1, 43	10.642	**0.002**
Picture number * spatial congruency * semantic congruency	1, 43	0.113	0.738	1, 43	0.208	0.65

Table 5.2 Results of repeated measures ANOVA: influence on picture identification accuracy and mean visibility rating of masked pictures.

Semantic congruency however, did affect picture identification accuracy (F and p-values are listed in Table 5.2, mean accuracies ± SEM are shown in Figure 5.2C). We also observed an effect of semantic correspondence on mean visibility rating (F and p-values are listed in Table 5.2, average ratings ± SEM are shown in Figure 5.2D) and a significant interaction between semantic and spatial congruency. Analysis of simple main effects showed that the effect of semantic congruency was significant for both collocated (F(1,43) = 19.2, p < 0.001) and disparate stimuli (F(1,43) = 7.159, p = 0.011), but we only observed a significant effect of spatial alignment for semantically congruent (F(1,43) = 9.108, p = 0.004), but not incongruent trials (F(1,43) = 0.44, p = 0.511).

Analysis with picture visibility used as factor

Not surprisingly, sound localization accuracy was overall affected by spatial alignment of the stimuli, but additionally we observed a main effect of image visibility and interactions between visibility x spatial congruency, visibility x semantic congruency and visibility x spatial x semantic congruency (F and p-values are listed in Table 5.3, mean accuracies ± SEM are shown in Figure 5.3A). We observed ventriloquist illusion (accuracy higher for collocated than disparate pairs) both for visible (simple main effect: F(1,33) = 21.857, p < 0.001) and invisible trials (simple main effect: F(1,33) = 12.133, p = 0.001). However, visibility affected sound localization accuracy in different ways for collocated and disparate stimuli. For spatially congruent pairs, observers performed better for visible trials (stronger multisensory enhancement in visible condition; F(1,33) = 7.674, p = 0.009), where for incongruent pairs, accuracy was higher for invisible comparing to visible trials (less ventriloquism in invisible condition; F(1,33) = 15.244, p < 0.001). Sound localization accuracy was influenced by semantic correspondences only in visible (F(1,33) = 4.530, p = 0.041), but not invisible trials (F(1,33) = 0.459, p = 0.503). As for the interaction of visibility x spatial x semantic congruency, simple effects of spatial alignment were all significant (all p < 0.005). Simple effect of visibility was not significant for collocated, semantically incongruent pairs (p = 0.459), but highly significant in all other conditions (p values <= 0.001). Simple effect of semantic correspondence was only significant for collocated stimuli, when picture was visible (p = 0.001).

Factor	Sound Localization Accuracy			Picture Identification Accuracy		
	df	F	p	df	F	p
Picture number	1, 33	3.972	0.055	1, 33	0.181	0.673
Spatial congruency	1, 33	21.896	**< 0.001**	1, 33	0	0.992
Visibility	1, 33	5.782	**0.022**	1, 33	60.074	**< 0.001**
Semantic congruency	1, 33	1.027	0.318	1, 33	34.283	**< 0.001**
Picture number * Spatial congruency	1, 33	1.864	0.181	1, 33	0.017	0.897
Picture number * Visibility	1, 33	0.089	0.767	1, 33	0.396	0.533
Spatial congruency * Visibility	1, 33	15.592	**< 0.001**	1, 33	0.202	0.656
Picture number * Spatial congruency * Visibility	1, 33	0.363	0.551	1, 33	0.35	0.558
Picture number * Semantic congruency	1, 33	0.845	0.365	1, 33	0.143	0.708
Spatial congruency * Semantic congruency	1, 33	0.817	0.372	1, 33	0.112	0.74
Picture number * Spatial congruency * Semantic congruency	1, 33	0.007	0.932	1, 33	0.039	0.845
Visibility * Semantic congruency	1, 33	4.705	**0.037**	1, 33	11.412	**0.002**
Picture number * Visibility * Semantic congruency	1, 33	0.893	0.351	1, 33	1.943	0.173
Spatial congruency * Visibility * Semantic congruency	1, 33	5	**0.032**	1, 33	1.411	0.243
Picture number * Spatial congruency * Visibility * Semantic congruency	1, 33	0.352	0.557	1, 33	0.405	0.529

Table 5.3 Results of repeated measures ANOVA: influence on sound localization and picture identification accuracies accounting for image visibility.

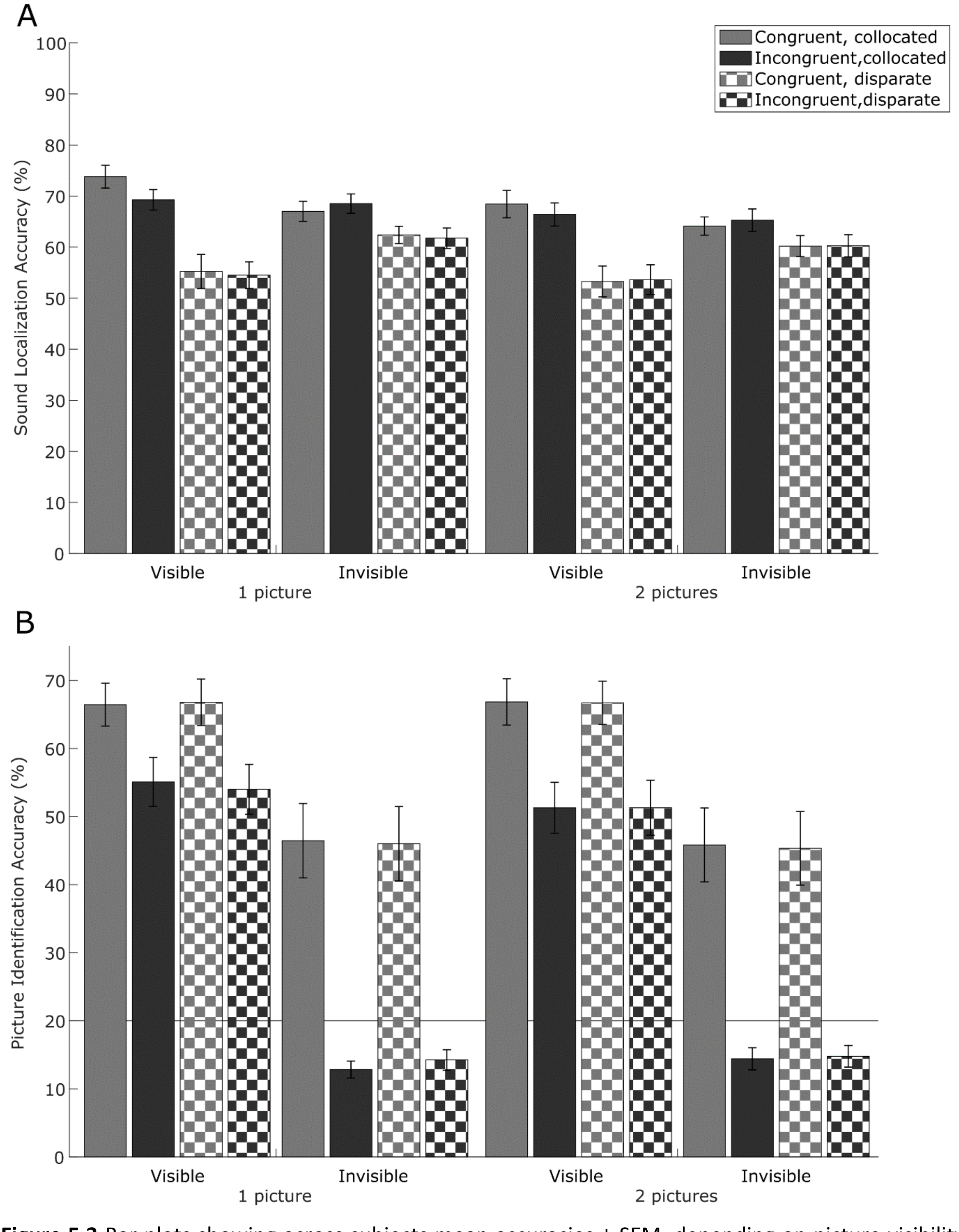

Figure 5.3 Bar plots showing across subjects mean accuracies ± SEM, depending on picture visibility (rating 0 was considered invisible, 1-3 visible). A. Sound localization accuracy (Left vs Right). B. Picture identification accuracy (choice of 1 out of 5).

Picture identification accuracy was only influenced by visibility and semantic congruency (F and p-values are listed in Table 5.3, mean accuracies ± SEM are shown in Figure 5.3B). As expected, accuracy for visible comparing to invisible trials was considerably higher. We also observed improved identification performance for semantically congruent pairs and an interaction between visibility and semantic congruency. Interestingly though, effects of semantic correspondence are significantly larger for invisible than visible conditions (paired t-test t(33) = 3.378, p = 0.002). As illustrated in Figure 5.3B, identification accuracy for invisible, semantically incongruent trials is below chance level, where for congruent ones it is close to accuracy in visible, incongruent condition. This suggests that the difference is driven by participants' response bias (responding with sound category when the picture is invisible), rather than an actual difference in perception.

Discussion

The first experiment confirmed that the ventriloquist illusion can be modulated by semantic congruency of audio-visual stimulus pairs. As we suspected, this modulation was significantly smaller for unilateral than bilateral presentations, which explains why it was not always observed in previous, low-powered studies, which used single picture presentations.

In the second experiment, employing forward-backward masking, observers' sound localization accuracy was still affected by spatial congruency (ventriloquist effect), but not by semantic correspondences between sound and picture. Importantly, the effect of semantic congruency was also obliterated for masked pictures that were judged visible. This suggests, that even for consciously perceived images, degraded visual processing of semantic features (i.e. in visual masking) is insufficient to influence multisensory integration. This result, along with studies showing McGurk illusion to falter under Continuous flash suppression (Ching et al., 2019; Palmer & Ramsey, 2012), supports the hypothesis that only low (i.e. spatial), but not high level (i.e. semantic) cues can affect multisensory integration in the absence of awareness.

Concurrently, two psychophysics studies have shown effects dependent on semantic congruency of unconscious, audio-visual signals (Faivre et al., 2014; Scott et al., 2018). While both crossmodal congruency priming (Faivre et al., 2014) and associative learning (Scott et al., 2018) for unconscious stimuli pose a challenge for GWT, neither in fact shows multisensory integration (i.e. there is no evidence for one modality affecting the other, written and spoken words do not form audio-visual objects (Bizley, Maddox, & Lee, 2016) and semantic associations can be created between pairs of stimuli within a single modality). Therefore, those interactions and ventriloquist/McGurk illusions could be driven by different mechanisms and rely on separate neural pathways. For example, it has been shown that nonconscious unimodal stimuli can enter the working memory and be maintained over several seconds (Dutta, Shah, Silvanto, & Soto, 2014; King et al., 2016; Pan, Lin, Zhao, & Soto, 2014). The mentioned effects could potentially result from comparison of the information stored in working memory and do not involve direct audio-visual communication.

Nevertheless, we observed significant effects of semantic correspondences on mean visibility rating and picture identification accuracy, which were higher for congruent stimulus pairs. These results replicate findings of previous studies showing supraliminal, semantically congruent sound affecting perception of visual stimuli that were obliterated from awareness. For instance, congruent sounds have been shown to reduce time for visual stimuli to reach awareness under Continuous Flash Suppression (Cox & Hong, 2015) and improve identification accuracy of masked pictures (Y.-C. Chen & Spence, 2010). Critically, these effects do not violate the assumptions of GWT and are not unique for multisensory signals. Familiarity of the image content itself can accelerate breaking Continuous Flash Suppression, which has been shown for recognizable words (Jiang et al., 2007) and for upright comparing to inverted images (Jiang et al., 2007; Stein et al., 2012). Similar facilitation in breaking CFS can be observed for visual stimuli matching the one in visual working memory (Gayet et al., 2013), which suggests that merely semantic priming could be enabling this faster access to perceptual awareness.

Our results show that sounds can boost masked pictures into participants' awareness depending on semantic congruency. Yet, invisible pictures do not bias perceived sound location depending on semantic correspondences. The latter finding suggests that semantic congruency between auditory and visual signals does not affect multisensory integration in the absence of awareness.

CHAPTER 6. CAUSAL METACOGNITION IN AUDIO-VISUAL PERCEPTION

Patrycja Delong, Uta Noppeney

Contributions:

PD and UN designed the study. PD acquired the data. PD analysed the data under supervision of UN. PD wrote the manuscript.

Abstract

Metacognition is the ability to evaluate the accuracy of one's own decisions. Metacognitive efficiency informs us about the extent to which observers can access the uncertainty of their perceptual representations. So far, the vast majority of research has been focused on the visual domain, while little is known about metacognitive processes for multisensory signals. It has been proposed that the mechanism underlying confidence judgements is shared across the sensory modalities (and perceptual domains); on the other hand, some studies argued that metacognition depends on early sensory representation. Those concepts though are not necessarily exclusive: structures responsible for metacognitive judgement that are shared across sensory modalities could utilize first order information related to individual senses. In the present psychophysics-EEG study, we investigated to what extent observers can access information about sensory uncertainties of integrated audio-visual signals, and whether multisensory binding cues (spatial and semantic congruency) affect their confidence about the causal structure of the event. Our results support claims that metacognition does generalize both across different tasks (spatial localization and unity judgement) and sensory modalities (auditory and audio-visual). We did not however, observe a relationship between confidence reports and neural sensory representations. Confidence judgements in audio-visual unity perception task suggest that observers only had access to their causal uncertainties, but not individual sensory uncertainties, when stimuli were integrated into a single percept. Interestingly, while both spatial and semantic congruency affected first order performance, only spatial, but not semantic congruency modulated subjects' confidence.

Introduction

Metacognition is the ability of observers to evaluate their own cognition. Metacognitive performance can be tested in so-called "type 2 task", in which subjects are asked how confident they are about their perceptual decisions in the "type 1 task" (Galvin et al., 2003). Type 1 task can refer to any perceptual choice e.g. discriminating between two signals, where type 2 task requires observer to provide introspective judgement about their task 1 performance. Metacognitive research has been heavily dominated by visual experiments (Faivre et al., 2017), with few studies employing other sensory modalities and even fewer multisensory signals. To our knowledge, only one study investigating confidence in a bimodal (audio-visual) task has been published so far. Faivre and colleagues compared metacognitive performance in unimodal and bimodal conditions and showed a correlation for respective metacognitive efficiencies, suggesting a common mechanism underlying metacognitive processes (Faivre et al., 2018). However, they specifically focused on a bimodal condition that did not involve audio-visual interactions (i.e. comparison, but not integration of auditory and visual input). Therefore, metacognition for integrated multisensory stimuli remains untrodden territory.

Multisensory integration is known to improve performance for multimodal comparing to unimodal tasks – a phenomenon called multisensory enhancement. If this improvement stems from reduction of sensory uncertainty (and assuming observers can monitor that uncertainty), multisensory integration should lead to an increase of confidence. An open question is how observers' confidence would change for multisensory signals, which provide conflicting perceptual estimates. For example, in the ventriloquist illusion, visual stimulus can influence spatial perception of sound, when they are presented in different locations (Jack & Thurlow, 1973). When observers experience ventriloquism, they report auditory and visual stimuli as originating from the same source, just like for audio-visual pairs presented in the same locations, creating perceptual metamers for physically different stimuli. In a recent paper, Deroy et. al (Deroy, Spence, et al., 2016) suggested that if observers can monitor their causal uncertainty, they still should be able to differentiate between perceptual metamers in the second order task, providing confidence judgements.

In the current psychophysics/EEG study, we investigated how low (spatial alignment) and high level (semantic congruency) combination cues affect uncertainty about causal structure of audio-visual events. We used highly reliable visual stimuli to i. ensure that observers would experience the ventriloquist illusion, ii. minimize the influence of visual sensory noise in the audio-visual condition. With auditory stimuli being the main source of sensory uncertainty in the audio-visual condition, uncertainty in this condition should be the same as in the unimodal auditory condition, unless additional factors play a role (i.e. multimodal interactions).

Crossmodal correspondences, such as semantic congruency, can facilitate binding of multisensory signals (for review see: Doehrmann & Naumer, 2008). In our recent study, we showed that semantic congruency can affect sound localization performance depending on spatial alignment of the stimuli (Chapter 5). Observers' performance was better for semantically congruent pairs, when they were presented in the same spatial location and worse when presented in two different locations. However, this semantic influence on audio-visual binding faltered for masked visual stimuli, even when they were judged as visible. This highlights the possibility that the effect of crossmodal correspondence could be merely a response bias rather than perceptual change. Previous experiments have shown that the ventriloquist illusion alters neural representations of sound location (Bonath et al., 2014, 2007, Chapter 4). If semantic correspondences can modulate this perceptual change in ventriloquism, we should be able to observe it at the neural level. We employed multivariate pattern analysis of EEG data to investigate whether decoding accuracy of audio-visual spatial unity and/or sound location depends on semantic congruency of the stimuli.

Accumulating evidence suggests that metacognitive processes have an underlying neural mechanism that is common for different cognitive tasks (for review see: Rouault, McWilliams, Allen, & Fleming, 2018) and for different sensory modalities (Ais et al., 2016; Beck et al., 2019; De Gardelle et al., 2016; Faivre et al., 2018). Metacognitive supramodality was shown for audition, vision, touch and warmth perception, but not for nociceptive pain, which authors suggested is due to unique affective aspects of pain (Beck et al., 2019). The prefrontal cortex has been proposed as a common "neural engine" for metacognition (Bang & Fleming, 2018; S. M. Fleming et al., 2012; Morales, Lau, & Fleming, 2018). At the same time, a few studies have shown early signatures of confidence, pointing towards dependence on sensory related information (Boldt & Yeung, 2015; Gherman & Philiastides, 2015; Zakrzewski, Wisniewski, Iyer, & Simpson, 2019). In the current study, we hypothesised that higher confidence would be related to more distinctive neural representations of type 1 task features (due to less noisy sensory estimates), which would result in improved decoding accuracy for the high confidence trials. We computed neural metacognitive sensitivity and efficiency using classifier predictions instead of subjects' responses to explore that.

Methods

Participants

After giving informed consent, 34 healthy young adults (27 females, 4 left-handed, mean age: 20.4 years, standard deviation: 2.1, range: 18-26) took part in the initial psychophysics experiment. 24 of those subjects participated (19 females, 4 left-handed, mean age: 19.9 years, standard deviation: 1.8, range: 18-26) in the subsequent EEG experiment. The study was performed in accordance with the principles outlined in the Declaration of Helsinki and was approved by the local ethics review board of the University of Birmingham.

Stimuli and apparatus

Participants sat in a dimly lit room in front of a computer monitor at a viewing distance of 95cm. Visual stimuli were two images (bird & dog) from the Bank of Standard Stimuli database (Brodeur et al., 2010, 2014). Images were displayed for 100ms on a grey background. On each trial, a square image (3.5 visual degree side size) was centred at ±2.5 visual angle along the azimuth from the middle of the screen.

Auditory stimuli were two sounds (bird & dog) downloaded from http:// www.findsounds.com (on 26/07/2017). The sounds were edited to start from the beginning of the sound file and to last for 150ms. Peak amplitudes of all the sounds were equalized with Audacity software (http://audacityteam.org).

Auditory stimuli were presented via four external speakers, placed above and below the monitor at 65 dB sound pressure level. Upper and lower speakers were aligned vertically and located 2.5° to the left and 2.5° to the right of the monitor's centre (i.e. aligned with the picture locations along the azimuth).

Psychophysical stimuli were generated and presented on a PC running Windows XP using the Psychtoolbox version 3.0.11 (Brainard, 1997) running on MATLAB R2018b (Mathworks, Natick, Massachusetts).

Visual stimuli were presented on a 30" LCD monitor with a resolution of 2560 x 1600 pixels at a frame rate of 60Hz (NVIDIA Quadro 600 graphics card). Auditory stimuli were digitized at a sampling rate of 44.8 kHz via an M-Audio Delta 1010LT sound card. Exact audiovisual onset timing was confirmed by recording visual and auditory signals concurrently with a photodiode and a microphone. Audio-visual stimuli pairs were always presented in synchrony.

Experimental Design

Participants performed 3 tasks: i. visual spatial localization (unimodal), ii. auditory spatial localization (unimodal), iii. spatial unity judgment (audio-visual). In unimodal spatial localization tasks, an image/audio recording was presented in left or right location (±2.5°). Bird and dog pictures/recordings were presented equally often.

In a spatial ventriloquist paradigm (audio-visual), participants were presented with an image in one of two locations: left or right (±2.5°) and semantically congruent or incongruent sound originating from left or right (±2.5°). Hence, the 2 x 2 factorial design manipulated (1) spatial alignment of the audio-visual pairs (collocated vs disparate) and (2) semantic correspondence between them (congruent, incongruent) (Figure 6.1A).

Each trial started with the presentation of the fixation dot for a duration of 1000 - 1250ms (Figure 6.1B). Next, either sound or image was presented in unimodal trials or both sound and image were presented in bimodal trials. Response cues were presented 1000ms after stimulus onset.

After each unimodal presentation (auditory or visual) subjects reported stimulus location (Left or Right) and their confidence on a continuous scale (ranging from very unsure to very sure) using the computer mouse. After each audio-visual presentation, participants reported their spatial unity perception (Same or Different audio-visual locations) together with confidence using the computer mouse (Figure 6.1B). Response cues were presented on the screen until a response was provided, but no longer than for 3 seconds. The next trial started after a minimum of 1.5s after response cues were first shown (i.e. if a response was provided earlier, a fixation cross was presented up until 1.5s has passed).

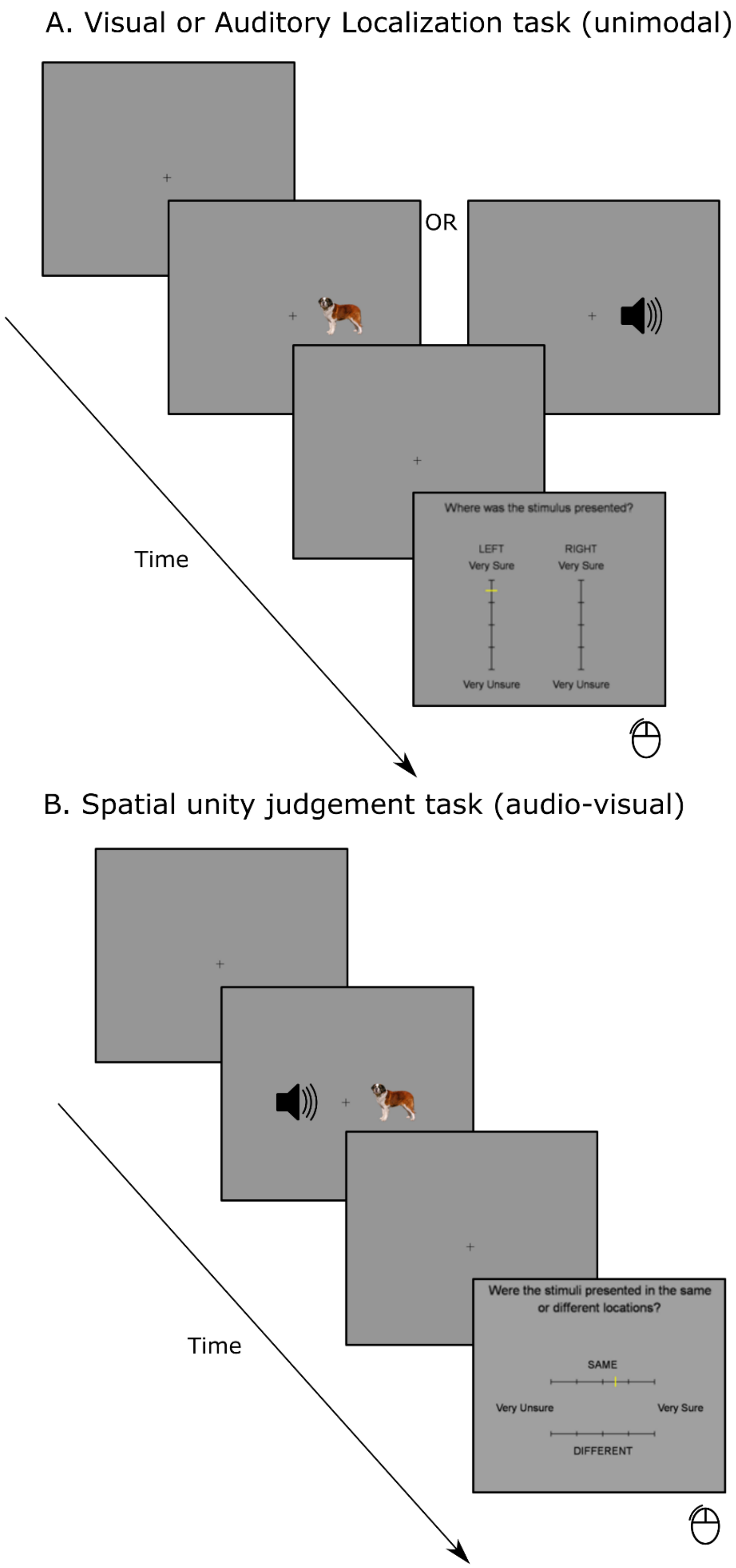
A. Visual or Auditory Localization task (unimodal)
OR
Where was the stimulus presented?
LEFT
Very Sure
RIGHT
Very Sure
Very Unsure
Very Unsure
Time
B. Spatial unity judgement task (audio-visual)
Were the stimuli presented in the same or different locations?
SAME
Very Unsure
Very Sure
DIFFERENT
Time

Figure 6.1 Example trial. A. Unimodal spatial localization task (visual or auditory). B. Multisensory spatial unity judgement task (audio-visual).

Experimental procedure and participant selection

The study included a one day psychophysics experiment (which served for participant selection for EEG study – see inclusion criteria below) and two day EEG experiment. All the results (behavioural and EEG analysis) are based on the full EEG experiment only.

During the one-day psychophysics study, participants performed i. unimodal auditory localization task and ii. spatial unity judgement task. The unimodal auditory task started with sound localization training, consisting of 32 trials, where subjects received feedback (Correct or Incorrect) after each response. After that, participants completed a further 64 trials (without feedback). Only subjects that achieved at least 70% sound localization accuracy were invited to take part in the EEG study. If a participant scored below 70% they repeated the task (including training). If their accuracy was below 70% the second time, they were excluded from the EEG study (1 participant with accuracy of 59% was excluded, 1 participant with accuracy of 69% was invited to continue). The purpose of this threshold was to ensure reliable sound localization within EEG subjects, which would hopefully correspond to relatively high decoding accuracy for the classifier trained on neural data. Moreover, reasonable sound localization accuracy is needed to discriminate between ventriloquism and no ventriloquism trials i.e. that shift in sound perception occurs and is not simply due to noisy auditory estimates.

Second task - audio-visual unity judgement, consisted of 192 trials (48 trials per each spatial x semantic congruency condition). To make sure that subjects understood the task and were reporting spatial unity and not semantic congruency, at least 70% accuracy (i.e. 70% Same responses) was required for collocated, but semantically incongruent trials. Secondly, to ensure sufficient numbers of trials with and without the ventriloquist illusion in the EEG study, we adopted a criterion of between 25 and 75% “same” responses for spatially disparate and semantically incongruent trials. If subject did not meet the above criteria, task was repeated; if the criteria were not met the second time, they were not invited to EEG experiment (this resulted in exclusion of 9 subjects).

The EEG experiment consisted of i. 4 runs of unimodal visual localization, ii. 4 runs of unimodal auditory localization and iii. 20 runs of audio-visual unity judgement split evenly over two testing days. Each block consisted of 96 trials, totalling 384 unimodal visual and 384 unimodal auditory trials (96 trials per semantic (dog, bird) x spatial (left, right) condition) and 1920 audio-visual trials (480 trials per semantic (congruent, incongruent) x spatial (collocated, disparate) condition). Block order was counterbalanced across participants. Position of ‘Same’ and ‘Different’ responses on the screen was swapped between first and second EEG sessions (top of bottom of the screen), similarly positions of ‘very sure’ and ‘very unsure’ confidence reports were flipped. This was a precaution to avoid any movement related confounds.

EEG data acquisition and preprocessing

Continuous EEG signals were recorded from 64 channels using Ag/AgCl active electrodes arranged in 10-20 layout (ActiCap, Brain Products GmbH, Gilching, Germany) at a sampling rate of 500Hz, referenced at FCz. Impedances were kept below 10kΩ. On each testing day 3D coordinates of EEG electrodes were digitized using Polhemus Fastrack (Polhemus Corp., Colchester, US).

Preprocessing was performed using MATLAB R2016b (Mathworks, Natick, Massachusetts) and Fieldtrip toolbox (Oostenveld et al., 2011). Muscle artefacts and noisy channels (0.4 channels on average) were identified based on visual inspection and rejected. Continuous EEG signals were high-pass filtered to 0.05 Hz and low-pass filtered to 50Hz. Independent component analysis (ICA) was applied to correct for eye movements and heartbeat artefacts. Eye blink and heartbeat-related components were identified based on visual inspection of component topographies and time-courses. Between 1 and 7 ICA components were removed (2.9 components on average). Noisy channels were interpolated using weighted average of neighbouring channels, based on sensor positions from Polhemus recordings. Data was segmented into -150:1000ms epochs relative to target stimulus (i.e. image and/or sound presentation) onset and re-referenced to average reference. After re-referencing, FCz electrode was appended, so that signals from 65 channels were entered into the analysis. Trials containing artefacts were rejected. Data was baseline corrected using 150ms prestimulus period and downsampled to 100Hz.

Data analysis

Data was analysed using MATLAB R2016b (Mathworks, Natick, Massachusetts) and JASP (Marsman & Wagenmakers, 2017). Significant interactions in ANOVAs were characterized by computing simple main effects for each factor. When analysis of simple effects did not explain the interaction (i.e. all simple main effects were significant and in the same direction), paired two-tailed t-tests were used to test for significant difference in size of those effects between conditions.

Behavioural analysis

Mean accuracy and confidence ratings were computed for each subject and modality (auditory, visual, audio-visual) and compared using a repeated measures ANOVA. If the sphericity assumption was violated, Greenhouse-Geisser correction was applied.

Confidence in unimodal and multimodal processing

Metacognitive measures (type 2 performance) can be influenced by task performance (type 1) (Galvin et al., 2003; Maniscalco & Lau, 2012). Highly reliable visual stimuli were deliberately used in this experiment to investigate visual influence on auditory processing (i.e. ventriloquist effect). Therefore, we expected observers' accuracy in visual trials to be close to 100% and strongly exceed the performance in auditory and audio-visual modalities, making it not comparable. This was indeed the case (details in the results section), therefore the analysis of metacognitive parameters was limited to auditory and audio-visual conditions. Metacognitive sensitivity and efficiency were computed as meta d' and meta d'/d' ratio respectively (Maniscalco & Lau, 2012), using MATLAB code available at: www.columbia.edu/~bsm2105/type2sdt/. Continuous confidence ratings (ranging from 0 to 1) were binarized for each subject and modality using a median split. Type 1 sensitivity (d'), metacognitive sensitivity (meta d') and metacognitive efficiency (meta d'/d') were compared between modalities using paired t-tests. In addition, we tested whether metacognitive efficiency for auditory and audio-visual tasks was correlated by computing a Pearson correlation coefficient.

Finally, task performance was compared for high/low confidence trials within each modality.

Audio-visual spatial unity

For audio-visual unity judgement task, confidence and accuracies were also calculated and compared between conditions using a 2 (spatial alignment: collocated, disparate) x 2 (semantic congruency: congruent, incongruent) factorial repeated measures ANOVA. A significant main effect of spatial alignment would provide evidence for the ventriloquist illusion, and a significant effect of semantic congruency would indicate it's influence on audio-visual binding. Additionally, rmANOVAs were used to i. compare confidence ratings between conditions depending on subjects' response (i.e. "same" or "different" locations judgements) ii. compare accuracies between conditions using confidence as a factor, alongside spatial and semantic congruency.

EEG analysis

Multivariate pattern analysis was used to investigate 1. the role of confidence in unimodal and multimodal processing of spatial information and 2. how visual stimulus affects auditory processing depending on spatial and semantic congruency.

Multivariate analyses were performed in MATLAB 2016B (Mathworks, Natick, Massachusetts) using the CoSMoMVPA toolbox (Oosterhof et al., 2016) and the Libsvm package (Chang & Lin, 2011). We trained a support vector machine (C=1) on single trial EEG activity patterns pertaining to i. the entire time window from 0 to 1000ms poststimulus or ii. 30ms sliding time windows (i.e. 65 channels x 3 time samples, for a 100Hz sampling rate). EEG signals were z-normalized in each channel with normalization parameters from training set being applied to testing set.

Using one sided t-tests we investigated whether decoding accuracy based on the entire time window was better than chance. For the sliding time window analysis we report p-values corrected for multiple comparisons using the Threshold Free Cluster Enhancement procedure (Smith & Nichols, 2009) with sign-permutation test (based on 10000 iterations) as implemented in CoSMoMVPA (Oosterhof et al., 2016). Unless stated otherwise, results are reported for $p < 0.05$.

Confidence in unimodal and multimodal processing

This analysis was again limited to auditory and audio-visual tasks, as subjects' confidence in visual trials was always high, so median split division of confidence ratings would introduce an artificial low confidence condition.

For the auditory spatial localization task, a support vector machine was trained to discriminate between Left and Right locations, where the number of bird and dog presentations was balanced within each training and testing fold. For the audio-visual unity judgement task, the classifier was trained to discriminate between Collocated and Disparate presentations, where the number of spatial combinations (for disparate: audio Left & visual Right, audio Right & visual Left; for collocated: audio & visual Left, audio & visual Right) and semantic combinations (audio bird & visual bird, audio dog & visual dog, audio bird & visual dog, audio dog & visual bird) was balanced within each fold. In both cases, classification was performed using 10 fold cross-validation.

D-primes, meta d-primes and meta d'/d' ratios were computed and analysed in the same fashion as in the behavioural analysis, but using classifier predictions instead of type 1 task responses. Additionally, we performed classification separately for high and low confidence conditions in both modalities to compare decoding accuracy.

Temporal dynamics of auditory and visual spatial representations

To investigate how similar auditory and visual spatial representations are over time, we performed temporal generalization analysis (King & Dehaene, 2014). For crossmodal generalization (i.e. from auditory to visual), we trained the classifier on balanced trials from one modality (balanced for locations and semantic type) and tested on trials from the other modality (again balanced for spatial and semantic conditions). For within modality time generalization, we have split balanced data into halves (each containing equal numbers of trials from each condition), where one half was used for training and one for testing.

Audio-visual spatial unity

Analysis of spatial processing in the audio-visual task was performed for i. auditory stimulus location (Left vs Right classification), and ii. unity of audio-visual stimuli (Collocated vs Disparate classification). Locations of the auditory stimuli were predicted using the classifier trained on unimodal auditory presentations and tested on the audio-visual task (only non-occipital EEG channels (55 channels) were used for the generalization to minimize confounds from visual stimuli). For auditory to audio-visual trial generalization, equal numbers of trials from each condition (2 sound files x 2 spatial location) were used in the training set, with all audio-visual trials used for testing. Decoding accuracy is given as balanced accuracy for all audio-visual conditions (i.e. mean accuracy based on all audio-visual conditions: 2 semantic congruency levels x 2 spatial alignment levels). Audio-visual spatial unity (collocated vs disparate predictions) was decoded from EEG activity using the unity classifier described earlier (‘Confidence in unimodal and multimodal processing’ section). A third classifier was trained to discriminate between semantically congruent and incongruent audio-visual pairs (balanced for all 4 semantic x 4 spatial combinations).

For the time window analysis, accuracy of i. spatial auditory decoding and ii. audio-visual spatial unity decoding was simply compared for spatial and semantic congruency conditions using a rmANOVA. We also performed analysis using participants’ response as a factor. This was only done for disparate audio-visual trials due to low numbers of ‘Different’ responses for collocated pairs. For disparate pairs, ‘same’ and ‘different’ responses equate to ventriloquist and non-ventriloquist trials. To investigate temporal dynamics of semantic, spatial and response related influence on decoding accuracy, we computed neural equivalents of effects that were detected at the behavioural level. The main effect of spatial alignment was calculated as (A stands for decoding accuracy):

$$E_{\text{Spatial alignment}} = \frac{\left(A_{\text{Collocated, Congruent}} + A_{\text{Collocated, Incongruent}} - A_{\text{Disparate, Congruent}} - A_{\text{Disparate, Incongruent}}\right)}{2}$$

The behavioural simple main effect of semantic congruency was only significant for disparate trials, and accuracy was higher for incongruent trials, so neural effect was computed as:

$$E_{\text{Semantic congruency}} = A_{\text{Disparate, Incongruent}} - A_{\text{Disparate, Congruent}}$$

The simple main effect of response was calculated as:

$$E_{Response,\, Disparate\ stimuli}$$

$$= \frac{\left(A_{Disparate,\, Incongruent, Different} + A_{Disparate,\, Congruent, Different} - A_{Disparate,\, Incongruent, Same} - A_{Disparate,\, Congruent, Same}\right)}{2}$$

We computed the above effects using classifier predictions for each timepoint. Statistical analysis was performed in the same way as for decoding accuracy, with the effects compared to a mean difference of zero.

Results

Behavioural results

Confidence in unimodal and multimodal processing

Participants' performance significantly differed for each modality (results with Greenhouse-Geisser correction: F(1.494,34.363) = 120.9, $p < 0.001$, post hoc statistics are listed in Table 6.1), with visual accuracy approaching 100% correct (see Figure 6.2). We also observed a significant effect of modality on mean confidence rating (F(2,46) = 47.3, $p < 0.001$, post hoc statistics are listed in Table 6.1) and post hoc tests showed that confidence in the visual modality was significantly higher than for the other two modalities, but there was no difference between confidence ratings for auditory and audio-visual modalities. This null effect was confirmed using Bayesian paired t-test (BF_{01} = 4.638). Metacognitive bias (i.e. overconfidence) should only be compared for tasks with identical performance, which is not the case here. However, as we have identical confidence ratings for auditory and audio-visual tasks, with significantly higher performance in the auditory task, we can conclude that subjects were overconfident in the audio-visual condition.

Post Hoc Comparisons – Accuracy					
Modalities		**Mean Difference**	**SE**	**t**	**p** $_{\text{bonf}}$
Auditory	**Visual**	-0.146	0.016	-8.833	**< .001**
Auditory	**Audio-visual**	0.129	0.013	9.639	**< .001**
Visual	**Audio-visual**	0.275	0.022	12.456	**< .001**

Post Hoc Comparisons – Confidence rating					
Modalities		**Mean Difference**	**SE**	**t**	**p** $_{\text{bonf}}$
Auditory	**Visual**	-0.188	0.023	-8.193	**< .001**
Auditory	**Audio-visual**	-0.002	0.022	-0.099	1.000
Visual	**Audio-visual**	0.185	0.021	8.809	**< .001**

Table 6.1 Post Hoc tests for accuracy and confidence reports for all sensory modalities.

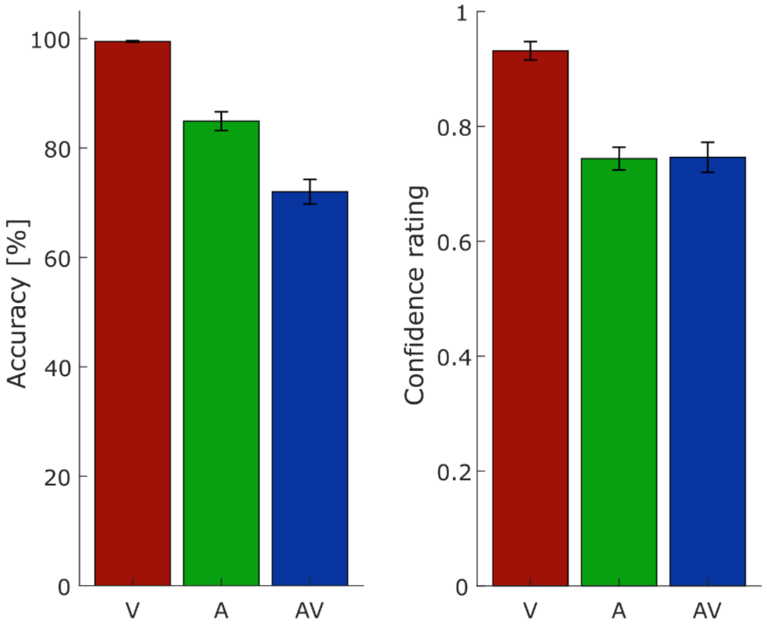

Figure 6.2 Bars showing accuracies and confidence ratings (across subject means ± SEM). Each bar represents different task/modality: V – visual localization task, A – auditory localization task, AV – audio-visual unity judgement task.

As expected, participants accuracy in the visual localization task was close to 100%, which yields it inappropriate for metacognitive analysis (i.e. in such a case it is not possible to accurately assess confidence on incorrect trials). Similarly, subjects' confidence rating was approaching the maximum value of one, therefore confidence analysis based on the median split (dividing data into two confidence bins (low vs high)) would introduce an artificial level of ‘low’ confidence, where subjective confidence rating was actually high. Because of this, confidence related analysis was limited to auditory and audio-visual conditions.

D-primes, metacognitive sensitivity and efficiency for auditory and audio-visual tasks are shown in the Figure 6.3. Both d-prime and metacognitive sensitivity ware significantly higher in the auditory localization task as indicated by two-tailed paired t-tests (for d’: t(23) = 10.048, p < 0.001; for meta d’: t(23) = 5.999, p < 0.001). Metacognitive efficiency did not differ between the tasks (t(23) = 1.824, p = 0.081), but the Bayes factor did not provide decisive evidence for the null hypothesis (BF_{01}= 1.12). We observed significant correlation of metacognitive efficiency between auditory and audio-visual tasks (Pearson’s R = 0.43, p = 0.026).

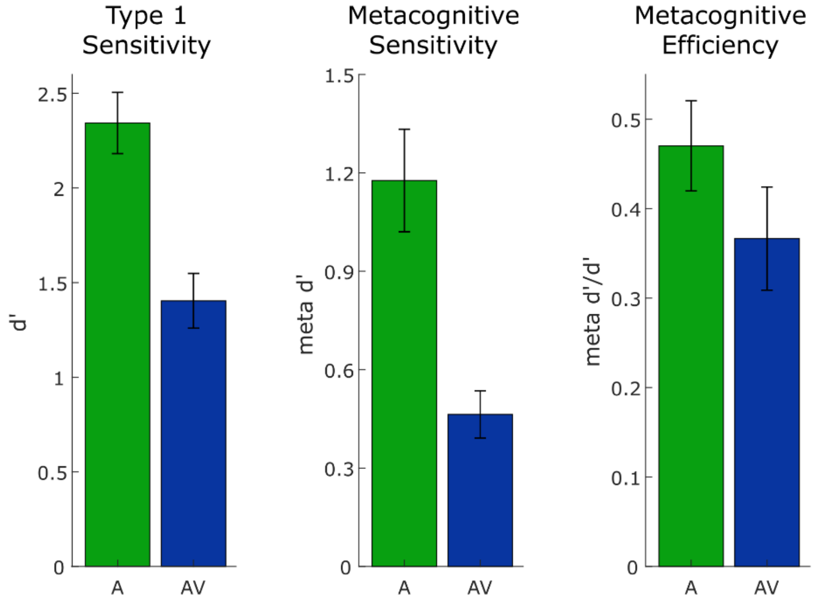

Figure 6.3 Bar plots showing first-order sensitivity (d’), metacognitive sensitivity (meta-d’), and metacognitive efficiency (meta-d’/d’) in auditory (A) and audio-visual (AV) modalities.

Audio-visual spatial unity

In the spatial unity judgment task (audio-visual), we observed a significant main effect of spatial alignment on accuracy (see detailed statistics in Table 6.2), which was lower for disparate than collocated stimuli (see Figure 6.4A). This indicates that subjects experienced the ventriloquist illusion, i.e. shift in sound perception toward visual stimulus, which leads to incorrect 'same location' judgement for disparate trials. Additionally, we observed a main effect of semantic congruency and an interaction of spatial and semantic congruency on participants' performance. Accuracy was higher for semantically congruent trials, when stimuli were collocated, but lower when they were presented in separate locations. Analysis of simple main effects showed that the effect of spatial alignment was significant for both semantically congruent (F(1,23) = 93.15, $p < 0.001$) and incongruent audio-visual pairs (F(1,23) = 96.08, $p < 0.001$), but we only observed a significant effect of semantic congruency for disparate (F(1,23) = 22.361, $p < 0.001$), but not collocated stimuli (F(1,23) = 3.03, p = 0.095). Interestingly, subjects' confidence was only affected by spatial alignment, but not semantic correspondence between stimuli (see Figure 6.4B; statistics in Table 6.2).

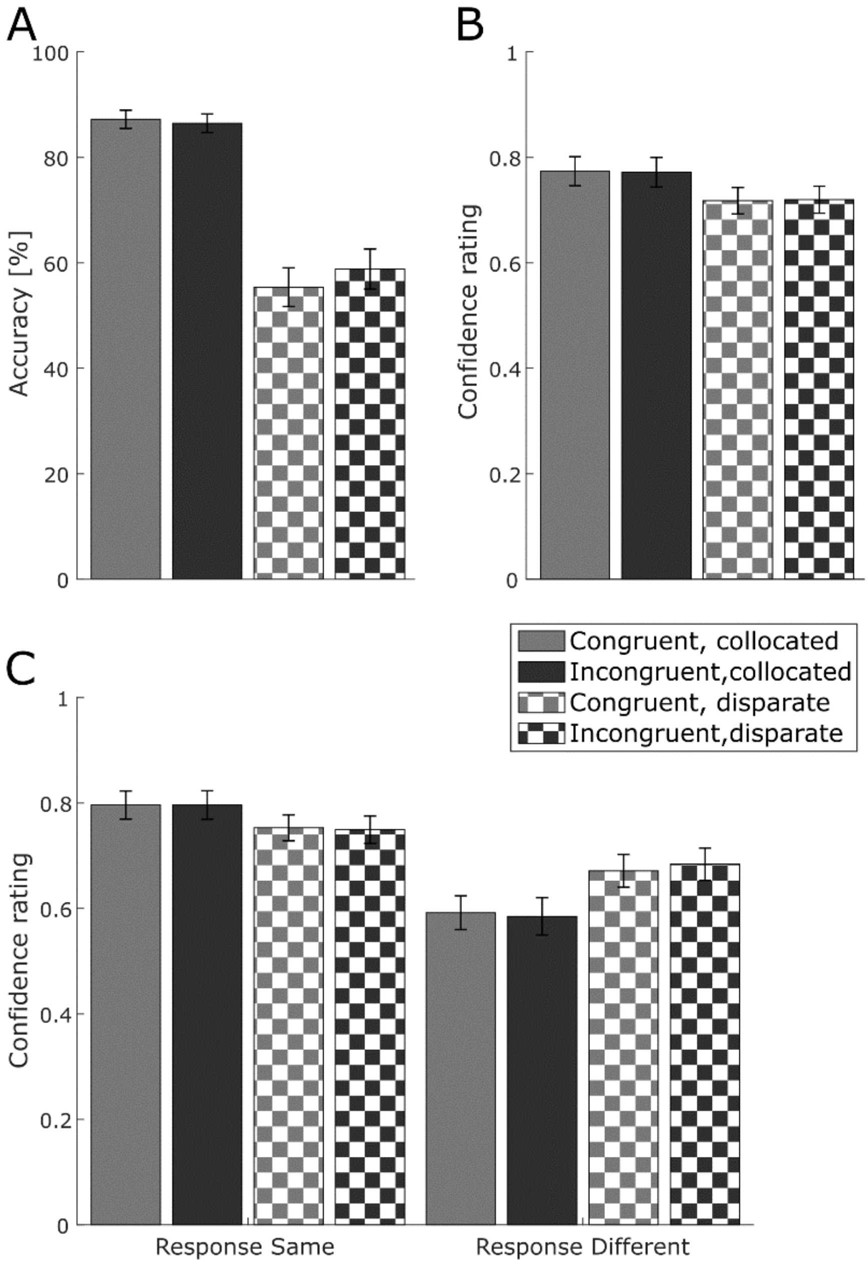

Figure 6.4 Accuracy and confidence in the audio-visual spatial unity judgement task. Bar plots showing across subject means ± SEM for different semantic and spatial combinations of audio-visual stimuli. A. Accuracy. B. Confidence ratings. C. Confidence ratings dependent on subjects' response.

Unity judgement Accuracy			
Factor	df	F	p
Semantic Congruency	1, 23	14.16	**0.001**
Spatial alignment	1, 23	95.19	**< 0.001**
Semantic Congruency * Spatial alignment	1, 23	19.10	**< 0.001**

Unity judgement Confidence			
Factor	df	F	p
Semantic Congruency	1, 23	0.003	0.959
Spatial alignment	1, 23	33.797	**< 0.001**
Semantic Congruency * Spatial alignment	1, 23	1.164	0.292

Table 6.2 Results of repeated measures ANOVA: influence on unity judgement accuracy and confidence.

Figure 6.4C illustrates mean confidence ratings for all conditions, including the unity report (same/different) as a factor. We observed a significant effect of spatial alignment, subjects' response, and their interaction (detailed statistics in Table 6.3). Participants' confidence was overall higher when they reported audio-visual coming from the same locations, whether this judgement was correct or not (i.e. confidence in 'Same' responses was higher than in Different' responses even for spatially disparate stimuli), with simple effects of response significant for both collocated (F(1,23) = 55.99, p < 0.001) and disparate stimulus pairs (F(1,23) = 18.7, p < 0.001). Simple effects of spatial alignment were significant for both 'same' (F(1,23) = 19.58, p < 0.001) and 'different' judgements (F(1,23) = 16.95, p < 0.001), but, unsurprisingly, have different directions (i.e. confidence is higher for disparate than collocated stimuli, when participants report them as coming from different locations and higher for collocated stimuli, when they are reported as coming from the same location).

Unity judgement Confidence with response as a factor			
Factor	df	F	p
Semantic Congruency	1, 23	0.018	0.895
Spatial alignment	1, 23	10.011	0.004
Response	1, 23	66.020	< 0.001
Semantic Congruency * Spatial alignment	1, 23	1.170	0.291
Semantic Congruency * Response	1, 23	0.293	0.593
Spatial alignment * Response	1, 23	18.947	< .001
Semantic Congruency * Spatial alignment * Response	1, 23	1.582	0.221

Table 6.3 Results of repeated measures ANOVA (with response as a factor): influence on unity judgement confidence.

Task performance for both sound localization and spatial unity judgement was affected by subjects' confidence (see Figure 6.5A). We observed significant effects of modality (see detailed statistics in Table 6.4), confidence and their interaction. All simple main effects were significant (modality for high (F = 9.432, p = 0.005) and low confidence (F = 212.053, p < 0.001); confidence in auditory (F = 79.22, p < 0.001) and audio-visual modalities (F = 15.81, p < 0.001)), with the simple main effect of confidence significantly higher in auditory modality (t(23) = 6.072, p < 0.001) and the simple main effect of modality was in higher confidence trials (t(23) = 6.072, p < 0.001).

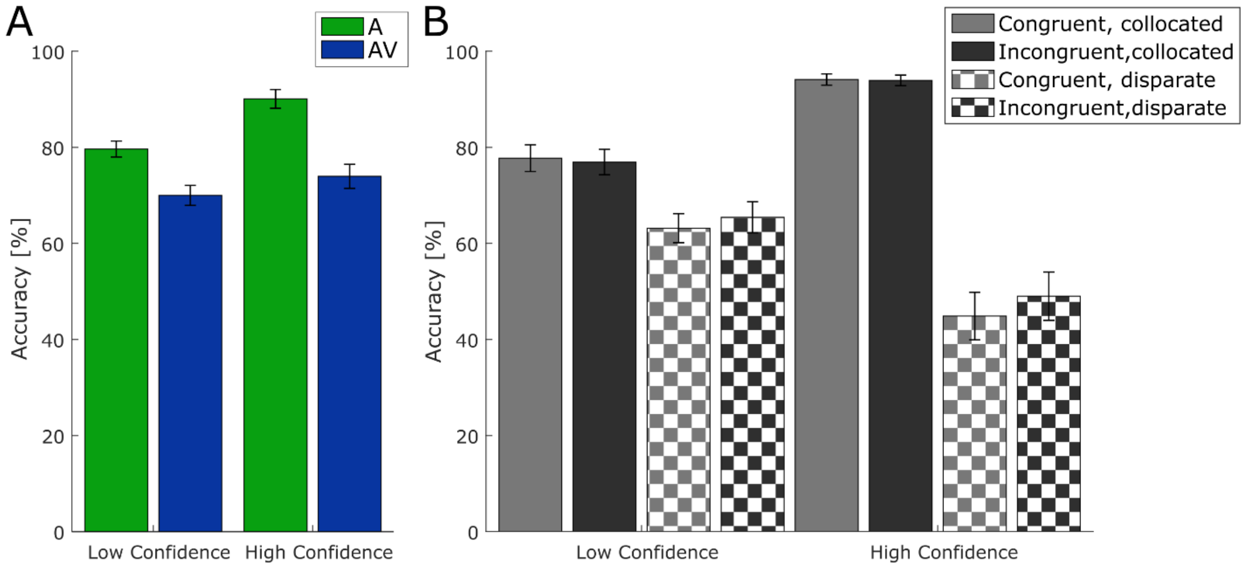

Figure 6.5 Performance depending on confidence ratings (across subject means ± SEM). A. Accuracy in auditory and audio-visual tasks. B. Accuracy in audio-visual task depending on stimuli properties (spatial alignment and semantic congruency).

Accuracy in Auditory and Audio-Visual tasks

Factor	df	F	p
Modality	1, 23	94.19	< 0.001
Confidence	1, 23	36.86	< 0.001
Modality * Confidence	1, 23	57.50	< 0.001

Accuracy in Audio-Visual task

Factor	df	F	p
Semantic Congruency	1, 23	13.731	0.001
Spatial alignment	1, 23	61.999	< .001
Confidence	1, 23	0.046	0.832
Semantic Congruency * Spatial alignment	1, 23	15.233	< .001
Semantic Congruency * Confidence	1, 23	3.212	0.086
Spatial Alignment * Confidence	1, 23	51.517	< .001
Semantic Congruency * Spatial alignment * Confidence	1, 23	1.502	0.233

Table 6.4 Results of repeated measures ANOVA (with response as a factor): influence on accuracy.

Analysis of audio-visual task performance with confidence as a factor (see Figure 6.5B) replicated the results of the basic analysis, i.e. main effects of spatial alignment, semantic congruency and their interaction (statistics listed in Table 6.4), but it did not yield a main effect of confidence. We did, however, observe an interaction between confidence and spatial alignment. Simple main effects of confidence were significant for both collocated (F = 57.08, p < 0.001) and disparate stimuli (F = 26.56, p < 0.001), but they had opposite directions, i.e. accuracy for collocated stimuli was higher for trials rated with high confidence, but for disparate stimuli, accuracy was higher for trials with low confidence ratings.

EEG results

Confidence in unimodal and multimodal processing

Neural d-primes were significantly different between auditory and audio-visual tasks (t(23) = 4.301, p < 0.001, two tailed test). There was no significant difference however, neither in metacognitive sensitivity (t(23) = 0.608, p = 0.549) with a Bayesian paired t-test confirming null hypothesis (BF_{01} = 3.937), nor in metacognitive efficiency (t(23) = 1.193, p = 0.245, BF_{01} = 2.473). We did not observe a significant correlation between neural metacognitive efficiency for auditory and audio-visual modalities (Pearson's R = -0.257, p = 0.226).

Neural d-primes for both auditory (t(23) = 7.335, p < 0.001, two tailed test) and audio-visual modalities (t(23) = 7.452, p < 0.001) were significantly greater than zero, which confirms better than chance performance of the classifier. Neural meta d' however, were not different from zero neither in auditory (t(23) = 0.934, p = 0.36, BF_{01} = 3.147), nor in audio-visual task (t(23) = 1.054, p = 0.303, BF_{01} = 2.832), which suggest no relationship between quality of relevant neural representations and confidence.

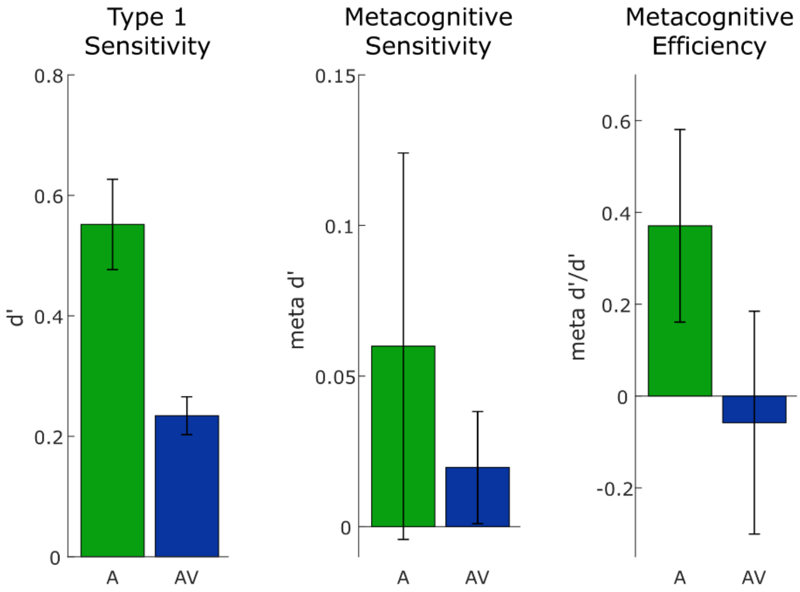

Figure 6.6 Bar plots showing first-order sensitivity (d'), metacognitive sensitivity (meta-d'), and metacognitive efficiency (meta-d'/d') computed based on classifier predictions for auditory (A) and audio-visual (AV) tasks.

To further investigate this, we performed the multivariate analysis separately for trials of high and low confidence (i.e. separate classifiers trained on either high or low confidence trials). If there is a difference in neural representation of spatial auditory position or audio-visual unity, this should be easier to detect if we train a classifier on high quality trials vs low quality trials (see Figure 6.7). Nevertheless, even using this method we did not observe a significant effect of confidence (F(1,23) = 0.213, p = 0.649). Unsurprisingly, we observed a significant effect of modality (F(1,23) = 6.354, p = 0.019).

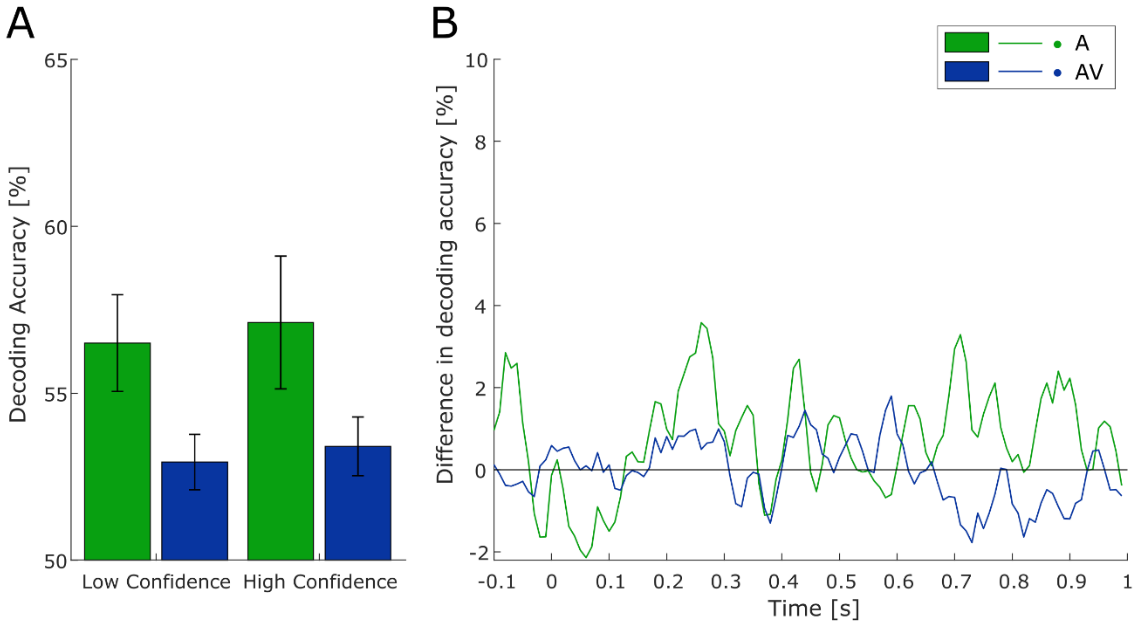

Figure 6.7 Classifier performance depending on confidence ratings (across subject means ± SEM) in auditory (A) and audio-visual (AV) tasks. SVM was trained to decode spatial location (Left vs Right) in auditory condition, and spatial alignment (Collocated vs Disparate) of stimuli pairs in audio-visual condition. A. Bar plots showing decoding accuracy for classifier using EEG data from entire time window from 0 to 1000ms. B. Difference in classifier performance between trials with high vs low confidence over time. Dots indicate difference significantly greater than zero.

Temporal dynamics of auditory and visual spatial representations

Spatial locations of visual stimuli could be successfully decoded from the EEG data from 60ms poststimulus onwards (see Figure 6.8). For auditory stimuli, decoding started only about 300ms poststimulus. At the same time, crossmodal generalization from auditory to visual modality is already significant at 210ms, and at 250ms for generalization from visual to auditory. This suggests that auditory representations must be available in the brain at earlier times. A recent study of Aller & Noppeney showed auditory spatial decoding already at 95ms poststimulus (Aller & Noppeney, 2019). This might be simply because they used much larger spatial disparities (i.e. 6 and 10 visual degree disparities, when we used only 5), which produced more distinctive neural representations. It had been shown that stronger underlying signals can be decoded earlier in time (Grootswagers et al., 2017).

Cross-temporal decoding within the visual modality is initially only decodable along the diagonal, but becomes more sustained over time after 200ms poststimulus. Temporal generalization for auditory and both crossmodal conditions is generally sustained over time, after 300 and 200ms poststimulus respectively.

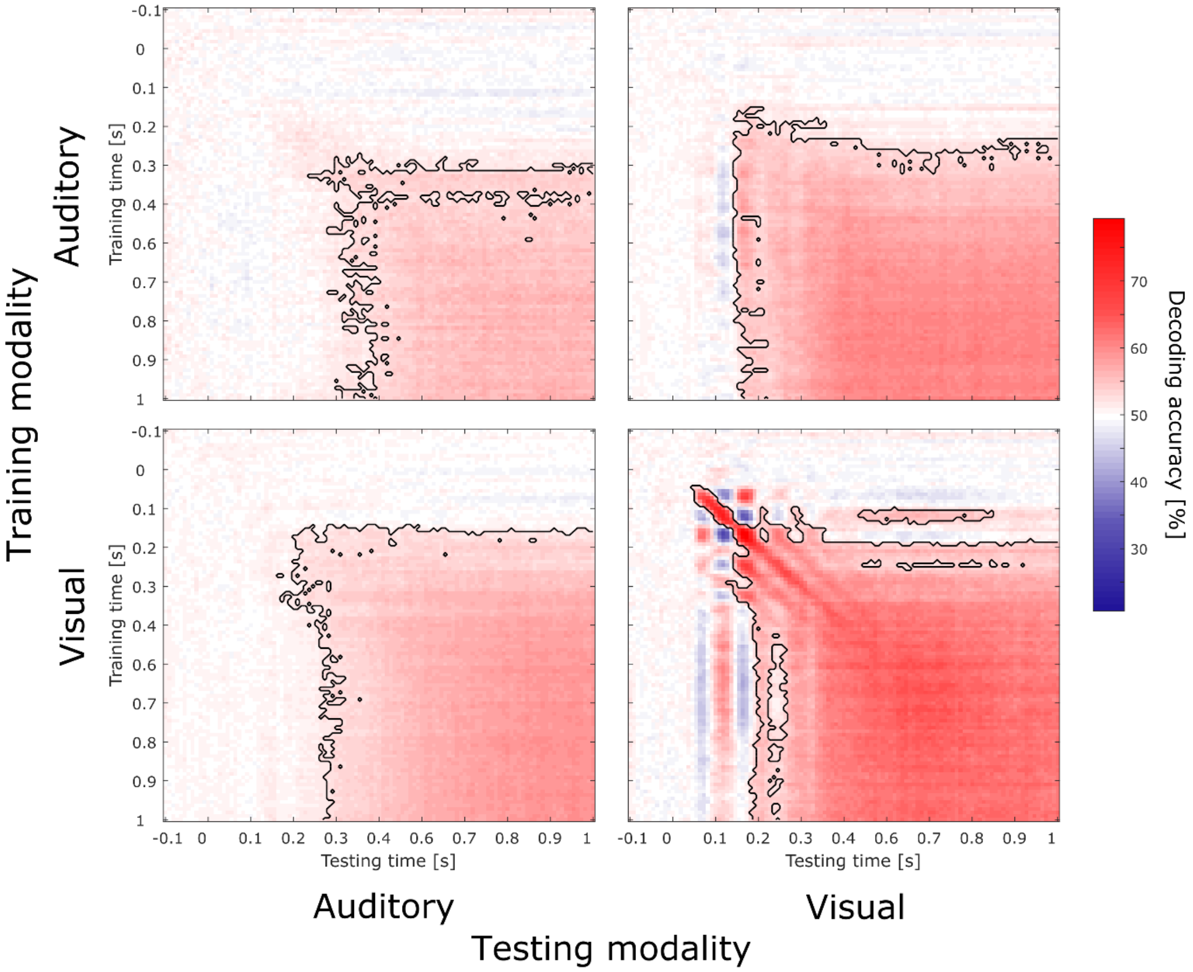

Figure 6.8 Temporal generalization within and across modalities for spatial localization of auditory and visual stimuli. Each matrix shows spatial decoding accuracy (Left vs Right) across training and testing times from -0.1 to 1 second. Black outline marks a decoding accuracy significantly better than chance (based on 2-dimensional spatial clustering).

Audio-visual spatial unity

For audio-visual trials, we were able to successfully decode spatial locations of auditory stimuli (t(23) = 6.055, p < 0.001) spatial unity of the stimuli (t(23) = 7.543, p < 0.001) and their semantic congruency (t(23) = 1.946, p = 0.032). Classifier performance over time is shown in Figure 6.9B. Significant decoding of auditory location started at 210ms post-stimulus and was maintained over time up to 1s. Better than chance classification of spatial alignment was first observed at 350ms post-stimulus and continued until the end of the analysed time window. Semantic correspondences could be decoded only within a 260 – 320ms time period and around 490ms, but not later.

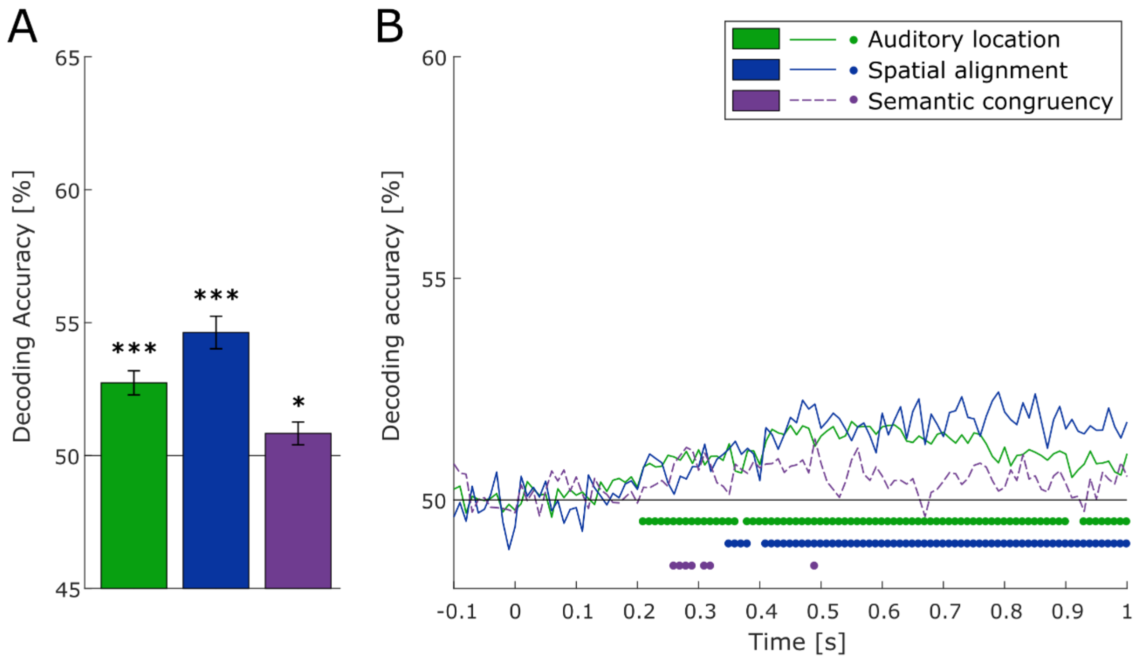

Figure 6.9 Accuracy of classifiers decoding i. auditory location, ii. spatial alignment and iii. semantic congruency, in audio-visual unity judgement task. A. Bar plots showing accuracy of classifiers using EEG data from entire time window from 0 to 1000ms. B. Classifier performance over time. Dots indicate performance significantly better than chance.

We observed a significant main effect of spatial alignment on decoding of auditory locations (see Figure 6.10A; statistics in Table 6.5), with decoding accuracy better for collocated than disparate stimulus pairs. The audio-visual classifier was trained to discriminate spatial alignment (collocated vs disparate) of stimulus pairs, therefore there should not be a difference in accuracy between discriminated conditions (numbers of trials from each condition was balanced for training, so accuracy should be optimized for both). As expected, we did not observe an effect on spatial alignment (see Figure 6.10D; statistics in Table 6.5). Semantic congruency did not affect performance either of auditory location classifier or of audio-visual unity classifier.

Analysis using participants' responses as a factor, showed a significant main effect of response on audio-visual unity, but not auditory location decoding accuracy (see Figure 6.10B&E; statistics in Table 6.6). As could be expected, classifier's unity predictions were more accurate for non-ventriloquist comparing to ventriloquist trials.

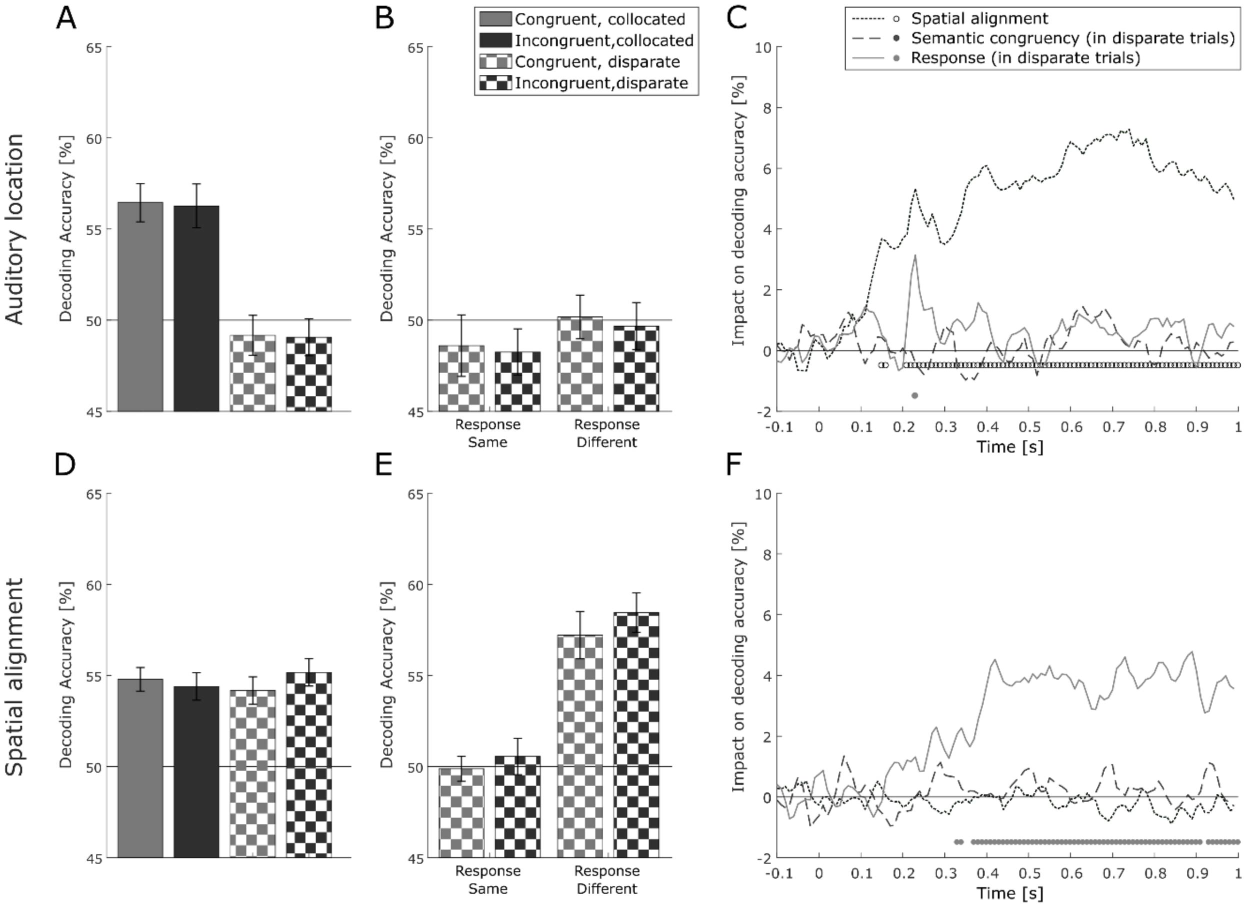

Figure 6.10 Classifier performance in the audio-visual task depending on audio-visual correspondences. A, B & C Results for auditory location classifier. D, E & F Results for spatial alignment classifier. A,B,D & E Bar plots showing decoding accuracy (across subjects mean ± SEM) based on entire time window from 0 to 1000ms. A & D Accuracy depending on spatial alignment & semantic congruency. B & E Accuracy for disparate trials depending on semantic congruency and subject's response. C & F Effects of i. spatial alignment, ii. semantic congruency and iii. subject's response on classification performance over time. Dots indicate effects significantly greater than zero. Temporal smoothing (average over 3 neighbouring data points – one on each side) was applied to plots C&F for visualization purposes (statistical analyses were performed for non-smoothed data).

Auditory predictions accuracy in audio-visual task

Factor	df	F	p
Semantic Congruency	1, 23	14.588	< 0.001
Spatial alignment	1, 23	0.174	0.680
Semantic Congruency * Spatial alignment	1, 23	0.005	0.944

Collocated vs Disparate Decoding Accuracy

Factor	df	F	p
Semantic Congruency	1, 23	0.032	0.859
Spatial alignment	1, 23	0.497	0.488
Semantic Congruency * Spatial alignment	1, 23	2.102	0.161

Table 6.5 Results of repeated measures ANOVA: influence on decoding accuracy.

Auditory predictions accuracy in audio-visual task for disparate condition

Factor	df	F	p
Semantic Congruency	1, 23	0.372	0.548
Response	1, 23	0.943	0.342
Semantic Congruency * Response	1, 23	0.009	0.924

Collocated vs Disparate Decoding Accuracy for disparate condition

Factor	df	F	p
Semantic Congruency	1, 23	2.490	0.128
Response	1, 23	37.208	< .001
Semantic Congruency * Response	1, 23	0.173	0.681

Table 6.6 Results of repeated measures ANOVA (with response used as a factor): influence on decoding accuracy.

The effect of spatial alignment on auditory location decoding was observed at 150ms post-stimulus and was maintained until the end of time window (see Figure 6.10C). Even though rmANOVA did not show a main effect of response on auditory prediction based on the entire time window, we observed significant influence of response on auditory location decoding around 220ms post-stimulus for sliding window analysis. The influence of response on audio-visual unity classification was detected at 320ms post-stimulus and continued until 1s (see Figure 6.10F).

Discussion

As expected, accuracy and confidence in the visual localization task were approaching 100%. Therefore, if there would be no interactions between the stimuli, auditory noise should be the only source of sensory noise in the audio-visual task. We observed significantly different d' primes for auditory and audio-visual tasks, which is a result of sensory interactions that are only present in the multimodal condition. Interestingly, mean confidence ratings were nearly identical for auditory and audio-visual tasks, despite significantly lower accuracy in the audio-visual condition. This shows metacognitive bias in the audio-visual task, i.e. observers were unaware of their poorer performance in the audio-visual condition, which led to overconfidence.

Differences in accuracy for unimodal and bimodal tasks can be easily explained by multisensory integration and the ventriloquist illusion. Lower accuracy for disparate than collocated stimuli confirms that the ventriloquist effect occurred. For disparate audio-visual trials, perception of sound was shifted towards the visual stimulus resulting in subjects incorrectly reporting stimulus pairs as originating from a single source. This discrepancy in task performance was reflected in participants' confidence ratings, which were lower for disparate stimuli. As for higher order crossmodal correspondences, we replicated the results of our previous study (Chapter 5) and observed significant influence of semantic congruency on audio-visual integration. Accuracy for collocated stimuli was higher for semantically congruent than incongruent pairs, and lower for semantically congruent than incongruent pairs when they were disparate. At the same time, semantic correspondences did not affect participants' confidence ratings. If semantic congruency affected the audio-visual binding tendency, it would modulate causal uncertainty, which should result in similar effects on mean confidence ratings as on accuracy.

Metacognitive sensitivity in the audio-visual task was significantly greater than zero, which confirms that participants were able to monitor their causal uncertainty (Figure 6.4A). They were also able to discriminate between perceptual metamers (the same type 1 task responses), which results in confidence ratings for collocated reports to be higher for collocated then disparate pairs, and for disparate reports higher for disparate than collocated pairs (Figure 6.4C). This result provides additional evidence that subjects had introspective access to their causal uncertainties. Interestingly however, they were more confident in correct than incorrect decisions for spatially collocated, but not disparate stimuli. For disparate stimuli, participants were actually more confident in ventriloquist (when disparate stimuli were perceived as coming from the same location - incorrect) than non-ventriloquist (correct) trials. This suggests that when audio-visual stimuli are integrated, observers no longer have access to unimodal sensory uncertainties. In contrast, recent study of the McGurk effect showed that subjects were less confident for McGurk stimuli presentations than both for conflicting and non-conflicting presentation (White et al., 2014). Note however, that this was a comparison of confidence for physically different stimulus pairs, and not illusion vs non-illusion trials, as it was in this study. The lack of introspective access to unimodal uncertainty does not explain why observers' confidence was lower in correct responses for disparate presentations (it would only explain lack of a difference). The reason for this could be that observers had high expectation of stimuli to be coming from a single source because they were presented in temporal synchrony, which led to perceptual conflict and increase of uncertainty for stimuli perceived as separate.

Metacognitive efficiencies for auditory and audio-visual tasks were correlated, which supports the view that metacognitive abilities generalize across different task (spatial localization and spatial unity judgement) and from unimodal to multimodal conditions. This is in line with previous findings suggesting supramodality of metacognition (De Gardelle et al., 2016; Faivre et al., 2018). The concept of a common underlying mechanism for metacognition across modalities does not exclude the possibility of sensory specific information influencing metacognition at early stages of neural processing, as was suggested in a few studies (Boldt & Yeung, 2015; Gherman & Philiastides, 2015; Zakrzewski et al., 2019). In the current experiment however, metacognitive sensitivity based on classifier predictions was at chance level for both auditory and audio-visual tasks (see Figure 6.6). Similarly, decoding accuracy did not differ between classifiers trained on high or low confidence trials. Together this suggests a lack of a relationship between the confidence judgement and quality of neural representation of first-order task features (i.e. spatial).

Spatial alignment affected decoding auditory location, but not spatial unity in audio-visual trials (Figure 6.10). This effect does not necessarily indicate perceptual change as it could also be the result of similarity of EEG topographies between auditory and visual spatial representations. Such similarity was shown to occur for spatial localization classifier trained on auditory and tested on visual stimuli (Aller & Noppeney, 2019) and we observed similar results here (Figure 6.8). Interestingly, the effect of spatial alignment emerges as early as 150ms poststimulus, where we only observed the crossmodal generalization significant 50ms later. In the Aller & Noppeney paper however, they showed crossmodal generalization already at 160ms, which suggests that this might be the source of spatial alignment effect on auditory location decoding.

We observed a significant effect of behavioural ventriloquism (unity response) on decoding accuracy of spatial alignment in disparate trials, both for classifier using 1s time window (Figure 6.10E) and for sliding window analysis over time. Effect of reported ventriloquism (response) becomes significant at around 320ms poststimulus, which is similar to our previous study showing modulation of neural ventriloquism by reported ventriloquism at 350ms (Chapter 4). Interestingly, behavioural ventriloquist illusion did not affect performance of auditory location classifier (trained on unimodal auditory condition) in time window analysis. We did, however, observe a small, significant effect at ~220ms poststimulus for analysis over time. This effect is much less pronounced than in our previous study (Chapter 4), which also used SVM trained on unimodal auditory locations. Discrepancy in results of the two studies could be due to different algorithms used (regression vs classification) and smaller audio-visual disparity in the present study (6 vs 5 visual degree), which could potentially lead to worse classifier performance. To investigate this we computed correlation coefficient between target and predicted locations for unimodal auditory condition in both studies, which showed that decoding performance was not significantly different between the two experiments (Fisher-transformed R in previous study R = 0.175 ± 0.03; in current study R = 0.222 ± 0.03; independent samples ttest t(40) = 1.074, p = 0.289, BF_{01} = 2.071). Therefore, difference in the effect of behavioural VE could not be the result of difference of classifier performance. The most likely explanation is difference in visual stimulation used in both experiments. In the previous study visual stimuli were shown under Continuous Flash Suppression, which results in weaker visual representations. Here, we observed spatial classification for visual stimuli at 92% (see Figure B.1, Appendix B), where it was only at 67% for visible stimuli and 53-55% for invisible stimuli presented under Continuous Flash

Suppression. As mentioned above EEG topographies of spatial visual and auditory locations are similar (as illustrated by crossmodal generalization in Figure 6.8), which can affect the performance of the auditory classifier in the audio-visual condition. This visual influence being weaker for suppressed visual stimuli would result in better performance of the auditory classifier. In fact, in the previous study effect of behavioural ventriloquism was smaller in visible than invisible trials, where visual decoding accuracy was significantly better. Taken together, these results support the idea that crossmodal generalization from auditory to audio-visual trials would have better performance when visual representations are less distinctive (i.e. under CFS, masking or judged invisible), which explains smaller response effects in the current study.

There was no effect of semantic congruency on reported confidence (Figure 6.4 B&C) nor decoding accuracy of auditory location and spatial unity (Figure 6.10), even though classifier could successfully discriminate between semantically congruent and incongruent pairs (Figure 6.9) and we did observe modulation of participants' performance. These results suggest that effects of semantic correspondences might be the result of a response bias, rather than perceptual change.

Appendix B

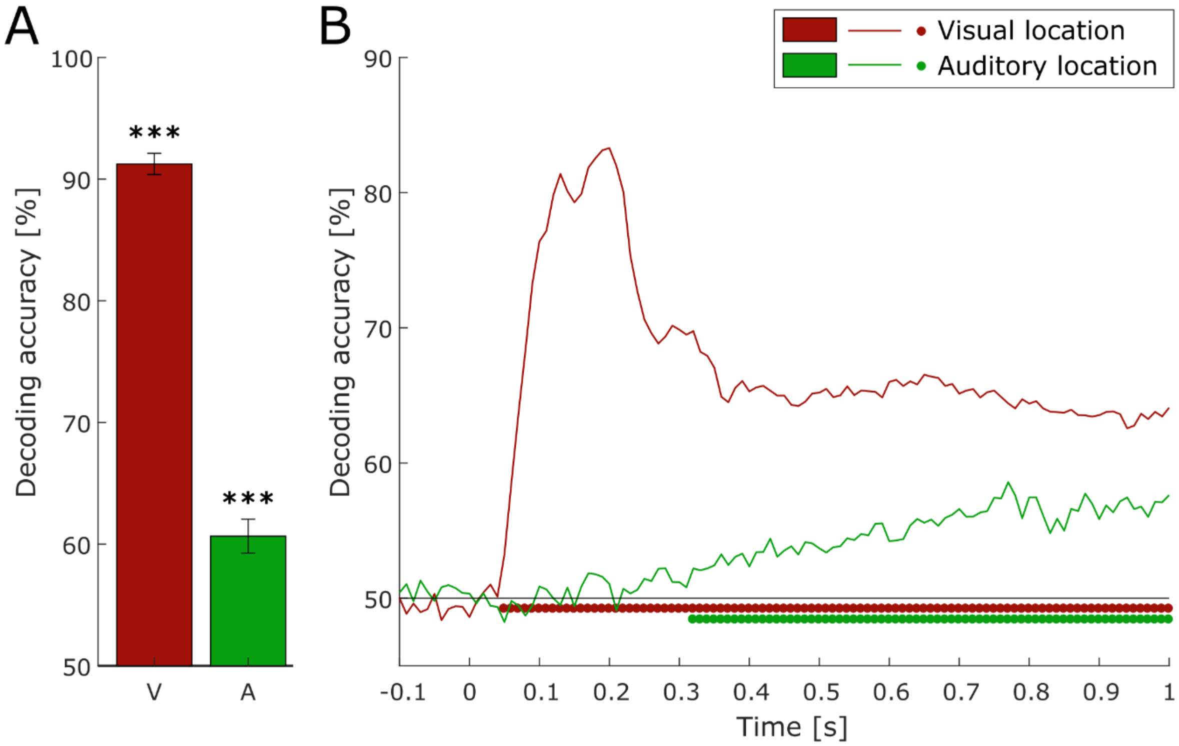

Figure B.1. Accuracy of classifiers decoding spatial locations (Left vs Right) in unimodal visual and auditory conditions. A. Bar plots showing decoding accuracy for classifier using EEG data from entire time window from 0 to 1000ms. B. Classifier performance over time. Dots indicate performance significantly better than chance.

CHAPTER 7. GENERAL DISCUSSION

The goal of this thesis was to understand the relationship of multisensory integration with subjective perceptual experiences and their related neural representations. I investigated how the process of creating unified multisensory objects can be affected by high and low-level binding cues, depending on observers' awareness and confidence about their sensory perception. Below I summarise the results of the experimental chapters and discuss their contribution to the field.

In the first studies I explored to what extent perception of supraliminal sounds can be affected by unseen visual stimuli. Across 4 experiments (Chapters 3-5), I consistently observed ventriloquist effect elicited by invisible visual signals. The effect was shown for two different types of audio-visual stimuli (abstract: beeps and flashes, and naturalistic: sound recordings and photographs) as well as two methods of visual suppression (Continuous Flash Suppression and backward-forward masking). While audio-visual integration was modulated by visual awareness, VE still occurred when visual signals were not seen. These results contradict Global Workspace Theory assumptions, showing that unaware signals from one sensory modality can travel through the global network and affect neural processing of signals in another sensory modality.

Two potential neural mechanisms that could mediate the "invisible ventriloquism" were proposed in Chapter 3: i. on VE trials invisible signals managed to escape the suppression and travelled through the visual hierarchy into higher-order association areas (such as Intraparietal Sulcus), where they were combined with auditory signals, ii. VE is dependent on direct connections between primary visual and auditory areas (Cappe & Barone, 2005; Falchier et al., 2002) and does not require higher-order processing. Difference in strength of audio-visual binding (indexed by VE) for visible and invisible flashes could be a result of either weaker visual spatial representations in the unseen condition, which do not modulate auditory perception as robustly (within the same neural pathway), or a similar initial mechanism irrespective of visibility, but additional higher level processing, only present for visible signals, that further modulates auditory perception. In the EEG experiment described in Chapter 4, decoding spatial location of invisible, incorrectly located flashes was only possible up until 300ms poststimulus, and the ventriloquist illusion nevertheless emerged in this condition. This suggests that this later spatial processing is not necessary for multisensory integration to occur.

The findings of Chapters 4 & 6 expand understanding of the neural basis of the ventriloquist illusion. Previous studies (Bonath et al., 2014, 2007) suggested that neural representations of illusion trials resemble sound representations of sounds presented in the location of visual stimuli. In Chapter 4, we computed neural ventriloquist as visual induced shift in sound location predicted by an algorithm trained on unimodal spatial representations of the sound (Equation 3 in Chapter 2). This allowed for direct quantification of the change in neural representations induced by ventriloquist effect. Importantly this neural ventriloquism was strongly dependent on the observers' reports, showing that perception of the illusion is caused by changes in brain activity. Similarly, in Chapter 6 predictions of audio-visual spatial unity were closely related to reported unity perception. Together with the fact that ventriloquism can be induced by unaware stimuli, this provides strong, corroborative evidence that the illusion reflects genuine change in perception, rather than a response bias.

In Chapters 5 & 6 I considered the role of semantic congruency in ventriloquism. While it is widely accepted that semantic correspondences can affect multisensory interactions, there is conflicting evidence regarding their influence on VE (e.g.: Colin et al., 2001 vs Kanaya & Yokosawa, 2011). Both experiments (Chapter 5 & 6) showed an increase in VE for semantically congruent comparing to incongruent stimuli pairs. This modulation was stronger for bilateral than unilateral presentations, which could explain discrepancies in previous findings mainly using unilateral presentations, where the effect could go undetected for small sample sizes. Interestingly however, I only observed the modulation for unmasked, but not masked stimuli, even if they were reported as visible. One explanation could be that unaware integration only supports low, but not high-level binding cues. Two experiments have failed to show McGurk effect for videos presented under CFS (Ching et al., 2019; Palmer & Ramsey, 2012), which supports this hypothesis. On the other hand, unaware crossmodal correspondences were successfully utilized for congruency priming (Faivre et al., 2014) and associative learning (Scott et al., 2018). Both approaches however, involve crossmodal interactions that could arise from stimulus comparison rather than integration (understood as merging of two signals into one perceptual event, generated by common cause). Therefore, these effects are likely to by mediated by different neural mechanisms and be governed by different rules.

If low and high-level features influence multisensory integration through different neural pathways, this could easily explain why only the Ventriloquist, but not McGurk effect or semantic modulation of VE, was observed for invisible stimuli. For instance, this could occur if only spatial information could travel through the direct connections between primary sensory cortices (which might be source of integration in the absence of awareness) and semantic representations influenced multisensory integration at later stage, within higher-order association cortices.

I was hoping that the EEG study in Chapter 6 would uncover the temporal dynamics of semantic modulation of ventriloquism and shed some light at the above hypothesis. Unfortunately, the effect of semantic correspondences was not observed at the neural level. This could be related to the fact that the influence of semantic congruency on VE was modest already at the behavioural level, and smaller than in the previous study. There are two differences between the studies that could potentially contribute to this decrease in the size of semantic modulation. First, subjects who frequently reported disparate locations for collocated, but semantically incongruent stimuli pairs were not invited to the EEG study, as it was not possible to determine whether they genuinely perceived the stimuli as coming from separate locations or simply misunderstood the task. That could lead to exclusion of subjects most prone to semantic modulation. This was not a concern in the experiment in Chapter 5, as participants were asked to report the sound location and not signals' spatial unity. Second, in neither of the studies semantic congruency was predictive of spatial alignment of the stimuli, but the EEG experiment consisted of much larger trial numbers and two testing days. Therefore, common source expectations could be decreasing over time, with second experiment affected more due to its duration. Future studies could try to investigate temporal dynamics of semantic modulation using bilateral presentations for which the influence is stronger.

Interestingly, even though I observed the behavioural effect of semantic congruency on unity judgement (Chapter 6), it did not influence confidence reports. Together with absence of the effect for masked stimuli (Chapter 5) and for neural spatial decoding (Chapter 6), this might suggest that the semantic modulation of VE is merely a response bias and does not influence perceived sound locations. The possibility that effects of semantic correspondences on audio-visual integration could stem from "postperceptual choices" was also considered in (Vatakis & Spence, 2008).

In the final empirical chapter, I looked at observers' causal uncertainty (uncertainty about causal structure of the event) in a ventriloquist paradigm. Results showed that audio-visual causal metacognitive efficiency is correlated with metacognitive efficiency in unimodal auditory localization task. This supports the notion that metacognitive performance generalizes across tasks and sensory modalities, including multimodal events (Ais et al., 2016; Beck et al., 2019; De Gardelle et al., 2016; Faivre et al., 2018; A. L. F. Lee, Ruby, Giles, & Lau, 2018; Morales et al., 2018). In contrast to previous findings, (Boldt & Yeung, 2015; Gherman & Philiastides, 2015; Zakrzewski et al., 2019), I did not find evidence for a relationship between confidence reports and reliability of neural sensory representations, indexed by classifier accuracy. Nevertheless, classifiers decoding accuracies were higher for high comparing to low confidence trials, even though the difference was not significant. Perhaps, the effect is small and should be put to test with a higher power study in the future.

Interestingly, psychophysics results of Chapter 6, showed that participants were able to metacognitively discriminate between correct and incorrect spatial unity reports, when audio-visual stimuli were spatially aligned (i.e. higher confidence for correct collocated responses), but not when they were displaced (higher confidence for incorrect collocated responses). This suggests that participants had limited access to their sensory uncertainties, when the stimuli were integrated. It seems that participants had high common cause expectations, as confidence was overall higher for “same location” comparing to “different locations” response. That could be related to the fact that stimuli pairs were always presented in temporal synchrony.

Observers were as confident in the audio-visual task as in the auditory localization task, despite reduced performance in the first one. This overconfidence in their spatial unity judgement performance, suggests that subjects were unaware of experiencing ventriloquist illusion. Together with studies showing VE induced by unseen visual stimuli (Chapters,3,4 &5) this provides strong evidence that multisensory integration in ventriloquist illusion in an automatic process.

REFERENCES

Adam, R., & Noppeney, U. (2014). A phonologically congruent sound boosts a visual target into perceptual awareness. *Frontiers in Integrative Neuroscience*, *8*(September), 70. https://doi.org/10.3389/fnint.2014.00070

Ais, J., Zylberberg, A., Barttfeld, P., & Sigman, M. (2016). Individual consistency in the accuracy and distribution of confidence judgments. *Cognition*, *146*, 377–386. https://doi.org/10.1016/j.cognition.2015.10.006

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/S0960-9822(04)00043-0

Aller, M., Giani, A., Conrad, V., Watanabe, M., & Noppeney, U. (2015). A spatially collocated sound thrusts a flash into awareness. *Frontiers in Integrative Neuroscience*, *9*(February), 1–8. https://doi.org/10.3389/fnint.2015.00016

Aller, M., & Noppeney, U. (2019). To integrate or not to integrate: Temporal dynamics of hierarchical Bayesian causal inference. *PLOS Biology*, *2*(3), 1–27. https://doi.org/10.1371/journal.pbio.3000210

Alsius, A., & Munhall, K. G. (2013). Detection of Audiovisual Speech Correspondences Without Visual Awareness. *Psychological Science*, *24*(4), 423–431. https://doi.org/10.1177/0956797612457378

Alves-Pinto, A., Baudoux, S., Palmer, A. R., & Sumner, C. J. (2010). Forward Masking Estimated by Signal Detection Theory Analysis of Neuronal Responses in Primary Auditory Cortex. *Journal of the Association for Research in Otolaryngology*, *11*(3), 477–494. https://doi.org/10.1007/s10162-010-0215-6

Arnold, D. H., Tear, M., Schindel, R., & Roseboom, W. (2010). Audio-visual speech cue combination. *PLoS ONE*, *5*(4). https://doi.org/10.1371/journal.pone.0010217

Baars, B. J. (1997). In the Theatre of Consciousness: The Workplace of the Mind. *Journal of Consciousness Studies*, *4*(4), 292–309. Retrieved from http://scholar.google.com/scholar?hl=en&btnG=Search&q=intitle:In+The+Theatre+of+Consciousness#3%5Cnhttp://scholar.google.com/scholar?hl=en&btnG=Search&q=intitle:In+the+Theatre+of+Consciousness:+The+Workplace+of+the+Mind%230

Baars, B. J. (2002). The conscious access hypothesis: origins and recent evidence. *Trends in Cognitive Sciences*, *6*(1), 47–52. https://doi.org/10.1016/S1364-6613(00)01819-2

Baars, B. J. (2005). Global workspace theory of consciousness: toward a cognitive neuroscience of human experience. *Progress in Brain Research*, *150*, 45–53. https://doi.org/10.1016/S0079-6123(05)50004-9

Bahrami, B., Vetter, P., Spolaore, E., Pagano, S., Butterworth, B., & Rees, G. (2010). Unconscious numerical priming despite interocular suppression. *Psychological Science*, *21*(2), 224–233. https://doi.org/10.1177/0956797609360664

Bang, D., & Fleming, S. M. (2018). Distinct encoding of decision confidence in human medial prefrontal cortex. *Proceedings of the National Academy of Sciences*, *115*(23), 201800795. https://doi.org/10.1073/pnas.1800795115

Barenholtz, E., Lewkowicz, D. J., Davidson, M., & Mavica, L. (2014). Categorical congruence facilitates multisensory associative learning. *Psychonomic Bulletin & Review*, *21*(5), 1346–1352. https://doi.org/10.3758/s13423-014-0612-7

Barrett, A. B., Dienes, Z., & Seth, A. K. (2013). Measures of metacognition on signal-detection theoretic models. *Psychological Methods*, *18*(4), 535–552. https://doi.org/10.1037/a0033268

Barutchu, A., Spence, C., & Humphreys, G. W. (2018). Multisensory enhancement elicited by unconscious visual stimuli. *Experimental Brain Research*, *236*(2), 409–417. https://doi.org/10.1007/s00221-017-5140-z

Beck, B., Peña-Vivas, V., Fleming, S., & Haggard, P. (2019). Metacognition across sensory modalities: Vision, warmth, and nociceptive pain. *Cognition*, *186*(February), 32–41. https://doi.org/10.1016/j.cognition.2019.01.018

Bedford, F. L. (1999). Keeping perception accurate. *Trends in Cognitive Sciences*, *3*(1), 4–11. https://doi.org/10.1016/S1364-6613(98)01266-2

Beierholm, U. R., Körding, K. P., Shams, L., & Ma, W. J. (2008). Comparing Bayesian models for multisensory cue combination without mandatory integration. *Advances in Neural Information Processing Systems 20*, *20*, 1–8. Retrieved from http://shamslab.psych.ucla.edu/publications/NIPS2008a.pdf

Békésy, G. v. (1947). A New Audiometer. *Acta Oto-Laryngologica*, *35*(5–6), 411–422. https://doi.org/10.3109/00016484709123756

Bertelson, P., & Aschersleben, G. (1998). Automatic visual bias of perceived auditory location. *Psychonomic Bulletin & Review*, *5*(3), 482–489.

Bertelson, P., & Aschersleben, G. (2003). Temporal ventriloquism: Crossmodal interaction on the time dimension---1. Evidence from auditory-visual temporal order judgment. *International Journal of Psychophysiology*, *50*(1--2), 147–155. https://doi.org/10.1016/S0167-8760(03)00130-2

Bertelson, P., Frissen, I., Vroomen, J., & de Gelder, B. (2006). The aftereffects of ventriloquism: Patterns of spatial generalization. *Perception & Psychophysics*, *68*(3), 428–436. https://doi.org/10.3758/BF03193687

Bertelson, P., Pavani, F., Ladavas, E., Vroomen, J., & de Gelder, B. (2000). Ventriloquism in patients with unilateral visual neglect. *Neuropsychologia*, *38*(12), 1634–1642. https://doi.org/10.1016/S0028-3932(00)00067-1

Bertelson, P., & Radeau, M. (1976). Ventriloquism, sensory interaction, and response bias: Remarks on the paper by Choe, Welch, Gilford, and Juola. *Perception & Psychophysics*, *19*(6), 531–535. https://doi.org/10.3758/BF03211222

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

Bertelson, P., Vroomen, J., de Gelder, B., & Driver, J. (2000). The ventriloquist effect does not depend on the direction of deliberate visual attention. *Perception & Psychophysics*, *62*(2), 321–332. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/10723211

Bertelson, P., Vroomen, J., Wiegeraad, G., & De Gelder, B. (1994). Exploring the relation between McGurk interference and ventriloquism. In *Proceedings of the Third International Congress on Spoken Language Processing, Yokohama, Japan, September 18-22, 1994.* (pp. 559–562). Baixas, France: International Speech Communication Association (ISCA). Retrieved from http://www.narcis.nl/publication/RecordID/oai:wo.uvt.nl:148924

Biderman, N., & Mudrik, L. (2018). Evidence for Implicit — But Not Unconscious — Processing of Object-Scene Relations. *Psychological Science*, *29*(2), 266–277. https://doi.org/10.1177/0956797617735745

Bizley, J. K., Maddox, R. K., & Lee, A. K. C. (2016). Defining Auditory-Visual Objects: Behavioral Tests and Physiological Mechanisms. *Trends in Neurosciences*, *39*(2), 74–85. https://doi.org/10.1016/j.tins.2015.12.007

Bizley, J. K., Nodal, F. R., Bajo, V. M., Nelken, I., & King, A. J. (2007). Physiological and anatomical evidence for multisensory interactions in auditory cortex. *Cerebral Cortex*, *17*(9), 2172–2189. https://doi.org/10.1093/cercor/bhl128

Björkman, M., Juslin, P., & Winman, A. (1993). Realism of confidence in sensory discrimination: The underconfidence phenomenon. *Perception & Psychophysics*, *54*(1), 75–81. https://doi.org/10.3758/BF03206939

Bode, S., & Haynes, J.-D. (2009). Decoding sequential stages of task preparation in the human brain. *NeuroImage*, *45*(2), 606–613. https://doi.org/10.1016/j.neuroimage.2008.11.031

Boldt, A., & Yeung, N. (2015). Shared neural markers of decision confidence and error detection. *Journal of Neuroscience*, *35*(8), 3478–3484. https://doi.org/10.1523/JNEUROSCI.0797-14.2015

Bonath, B., Noesselt, T., Krauel, K., Tyll, S., Tempelmann, C., & Hillyard, S. a. (2014). Audio-visual synchrony modulates the ventriloquist illusion and its neural/spatial representation in the auditory cortex. *NeuroImage*, *98*, 425–434. https://doi.org/10.1016/j.neuroimage.2014.04.077

Bonath, B., Noesselt, T., Martinez, A., Mishra, J., Schwiecker, K., Heinze, H.-J. J., & Hillyard, S. A. (2007). Neural basis of the ventriloquist illusion. *Current Biology*, *17*(19), 1697–1703. https://doi.org/10.1016/j.cub.2007.08.050

Brainard, D. H. (1997). The Psychophysics Toolbox. *Spatial Vision*, *10*(4), 433–436. https://doi.org/10.1163/156856897X00357

Breitmeyer, B. G. (2015). Psychophysical “blinding” methods reveal a functional hierarchy of unconscious visual processing. *Consciousness and Cognition*, *35*, 234–250. https://doi.org/10.1016/j.concog.2015.01.012

Brodeur, M. B., Dionne-Dostie, E., Montreuil, T., & Lepage, M. (2010). The Bank of Standardized Stimuli (BOSS), a New Set of 480 Normative Photos of Objects to Be Used as Visual Stimuli in Cognitive Research. *PLoS ONE*, *5*(5), e10773. https://doi.org/10.1371/journal.pone.0010773

Brodeur, M. B., Guérard, K., & Bouras, M. (2014). Bank of Standardized Stimuli (BOSS) Phase II: 930 New Normative Photos. *PLoS ONE*, *9*(9), e106953. https://doi.org/10.1371/journal.pone.0106953

Bruce, P., & Bruce, A. (2017). *Practical Statistics for Data Scientists*. *Practice*. Sebastopol: O'Reilly Media.

Burr, D., Banks, M. S., & Morrone, M. C. (2009). Auditory dominance over vision in the perception of interval duration. *Experimental Brain Research*, *198*(1), 49–57. https://doi.org/10.1007/s00221-009-1933-z

Cappe, C., & Barone, P. (2005). Heteromodal connections supporting multisensory integration at low levels of cortical processing in the monkey. *European Journal of Neuroscience*, *22*(11), 2886–2902. https://doi.org/10.1111/j.1460-9568.2005.04462.x

Cappe, C., Morel, A., Barone, P., & Rouiller, E. M. (2009). The thalamocortical projection systems in primate: an anatomical support for multisensory and sensorimotor interplay. *Cerebral Cortex*, *19*(9), 2025–2037. https://doi.org/10.1093/cercor/bhn228

Cappe, C., Thut, G., Romei, V., & Murray, M. M. (2010). Auditory-Visual Multisensory Interactions in Humans: Timing, Topography, Directionality, and Sources. *Journal of Neuroscience*, *30*(38), 12572–12580. https://doi.org/10.1523/jneurosci.1099-10.2010

Chang, C.-C., & Lin, C.-J. (2011). Libsvm. *ACM Transactions on Intelligent Systems and Technology*, *2*(3), 1–27. https://doi.org/10.1145/1961189.1961199

Charbonneau, G., Véronneau, M., Boudrias-Fournier, C., Lepore, F., & Collignon, O. (2013). The ventriloquist in periphery : Impact of eccentricity-related reliability on audio-visual localization. *Journal of Vision*, *13*(12), 20. https://doi.org/10.1167/13.12.20.doi

Chen, L., & Vroomen, J. (2013). Intersensory binding across space and time: a tutorial review. *Attention, Perception & Psychophysics*, *75*(5), 790–811. https://doi.org/10.3758/s13414-013-0475-4

Chen, Y.-C., & Spence, C. (2010). When hearing the bark helps to identify the dog: Semantically-congruent sounds modulate the identification of masked pictures. *Cognition*, *114*(3), 389–404. https://doi.org/10.1016/j.cognition.2009.10.012

Chen, Y.-C., Yeh, S.-L., & Spence, C. (2011). Crossmodal constraints on human perceptual awareness: Auditory semantic modulation of binocular rivalry. *Frontiers in Psychology*, *2*, 212. https://doi.org/10.3389/fpsyg.2011.00212

Chen, Z., & Saunders, J. a. (2016). Automatic adjustments toward unseen visual targets during grasping movements. *Experimental Brain Research*, *234*(7), 2091–2103. https://doi.org/10.1007/s00221-016-4613-9

Ching, A. S. M., Kim, J., & Davis, C. (2019). Auditory–visual integration during nonconscious perception. *Cortex*, *117*, 1–15. https://doi.org/10.1016/j.cortex.2019.02.014

Choe, C. S., Welch, R. B., Gilford, R. M., & Juola, J. F. (1975). The "ventriloquist effect": Visual dominance or response bias? *Perception & Psychophysics*, *18*(1), 55–60. https://doi.org/10.3758/BF03199367

Chong, S. C., Tadin, D., & Blake, R. (2005). Endogenous attention prolongs dominance durations in binocular rivalry. *Journal of Vision*, *5*(11), 1004–1012. https://doi.org/10.1167/5.11.6

Colin, C., Radeau, M., Deltenre, P., & Morais, J. (2001). Rules of intersensory integration in spatial scene analysis and speechreading. *Psychologica Belgica*, *41*, 131–144.

Colonius, H., & Diederich, A. (2004). Multisensory Interaction in Saccadic Reaction Time: A Time-Window-of-Integration Model. *Journal of Cognitive Neuroscience*, *16*(6), 1000–1009. https://doi.org/10.1162/0898929041502733

Conrad, V., Bartels, A., Kleiner, M., & Noppeney, U. (2010). Audiovisual interactions in binocular rivalry. *Journal of Vision*, *10*(10), 27. https://doi.org/10.1167/10.10.27

Conrad, V., Kleiner, M., Bartels, A., Hartcher O'Brien, J., Bülthoff, H. H., & Noppeney, U. (2013). Naturalistic stimulus structure determines the integration of audiovisual looming signals in binocular rivalry. *PLoS ONE*, *8*(8), e70710. https://doi.org/10.1371/journal.pone.0070710

Conrad, V., Vitello, M. P., & Noppeney, U. (2012). Interactions between apparent motion rivalry in vision and touch. *Psychological Science*, *23*(8), 940–948. https://doi.org/10.1177/0956797612438735

Cornsweet, T. N. (1962). The Staircase-Method in Psychophysics. *The American Journal of Psychology*, *75*(3), 485. https://doi.org/10.2307/1419876

Cox, D., & Hong, S. W. (2015). Semantic-based crossmodal processing during visual suppression. *Frontiers in Psychology*, *6*(June), 722. https://doi.org/10.3389/fpsyg.2015.00722

Crick, F., & Koch, C. (1990). Towards a neurobiological theory of consciousness. *Seminars in the Neurosciences*, *2*, 263–275.

Dahl, C. D., Logothetis, N. K., & Kayser, C. (2009). Spatial Organization of Multisensory Responses in Temporal Association Cortex. *Journal of Neuroscience*, *29*(38), 11924–11932. https://doi.org/10.1523/JNEUROSCI.3437-09.2009

De Gardelle, V., Le Corre, F., & Mamassian, P. (2016). Confidence as a common currency between vision and audition. *PLoS ONE*, *11*(1). https://doi.org/10.1371/journal.pone.0147901

Dehaene, S., & Changeux, J.-P. (2011). Experimental and Theoretical Approaches to Conscious

Processing. *Neuron*, *70*(2), 200–227. https://doi.org/10.1016/j.neuron.2011.03.018

Dehaene, S., & Naccache, L. (2001). Towards a cognitive neuroscience of consciousness: basic evidence and a workspace framework. *Cognition*, *79*(1–2), 1–37. https://doi.org/10.1016/S0010-0277(00)00123-2

Delong, P., Aller, M., Giani, A. S., Rohe, T., Conrad, V., Watanabe, M., & Noppeney, U. (2018). Invisible Flashes Alter Perceived Sound Location. *Scientific Reports*, *8*(1), 12376. https://doi.org/10.1038/s41598-018-30773-3

Deroy, O., Chen, Y.-C., & Spence, C. (2014). Multisensory constraints on awareness. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *369*(1641), 20130207. https://doi.org/10.1098/rstb.2013.0207

Deroy, O., Faivre, N., Lunghi, C., Spence, C., Aller, M., & Noppeney, U. (2016). The complex interplay between multisensory integration and perceptual awareness. *Multisensory Research*, *29*, 585–606. https://doi.org/10.1163/22134808-00002529

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

Diaconescu, A. O., Alain, C., & McIntosh, A. R. (2011). The co-occurrence of multisensory facilitation and cross-modal conflict in the human brain. *Journal of Neurophysiology*, *106*(6), 2896–2909. https://doi.org/10.1152/jn.00303.2011

Diederich, A., & Colonius, H. (2004). Bimodal and trimodal multisensory enhancement: Effects of stimulus onset and intensity on reaction time. *Perception and Psychophysics*, *66*(8), 1388–1404. https://doi.org/10.3758/BF03195006

Dixon, W. J., & Mood, A. M. (1948). A Method for Obtaining and Analyzing Sensitivity Data. *Journal of the American Statistical Association*, *43*(241), 109–126. https://doi.org/10.1080/01621459.1948.10483254

Dobreva, M. S., O'Neill, W. E., & Paige, G. D. (2012). Influence of age, spatial memory, and ocular fixation on localization of auditory, visual, and bimodal targets by human subjects. *Experimental Brain Research*, *223*(4), 441–455. https://doi.org/10.1007/s00221-012-3270-x

Doehrmann, O., & Naumer, M. J. (2008). Semantics and the multisensory brain: How meaning modulates processes of audio-visual integration. *Brain Research*, *1242*, 136–150. https://doi.org/10.1016/j.brainres.2008.03.071

Dutta, A., Shah, K., Silvanto, J., & Soto, D. (2014). Neural basis of non-conscious visual working memory. *NeuroImage*, *91*, 336–343. https://doi.org/10.1016/j.neuroimage.2014.01.016

Ehrenstein, W. H., & Ehrenstein, A. (1999). Psychophysical Methods. *Modern Techniques in Neuroscience Research*, 1211–1241. https://doi.org/10.1007/978-3-642-58552-4_43

Eo, K., Cha, O., Chong, S. C., & Kang, M.-S. (2016). Less Is More: Semantic Information Survives Interocular Suppression When Attention Is Diverted. *Journal of Neuroscience*, *36*(20), 5489–5497. https://doi.org/10.1523/JNEUROSCI.3018-15.2016

Ernst, M. O., & Banks, M. S. (2002). Humans integrate visual and haptic information in a statistically optimal fashion. *Nature*, *415*(6870), 429–433. https://doi.org/10.1038/415429a

Ernst, M. O., & Bülthoff, 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

Faisal, A. A., Selen, L. P. J., & Wolpert, D. M. (2008). Noise in the nervous system. *Nature Reviews. Neuroscience*, *9*(4), 292–303. https://doi.org/10.1038/nrn2258

Faivre, N., Arzi, A., Lunghi, C., & Salomon, R. (2017). Consciousness is more than meets the eye: a call for a multisensory study of subjective experience†. *Neuroscience of Consciousness*, *3*(1), 1–8. https://doi.org/10.1093/nc/nix003

Faivre, N., Berthet, V., & Kouider, S. (2012). Nonconscious influences from emotional faces: A comparison of visual crowding, masking, and continuous flash suppression. *Frontiers in Psychology*, *3*(MAY), 1–13. https://doi.org/10.3389/fpsyg.2012.00129

Faivre, N., Dubois, J., Schwartz, N., & Mudrik, L. (2019). Imaging object-scene relations processing in visible and invisible natural scenes. *Scientific Reports*, *9*(1), 1–13. https://doi.org/10.1038/s41598-019-38654-z

Faivre, N., Filevich, E., Solovey, G., Kühn, S., & Blanke, O. (2018). Behavioral, Modeling, and Electrophysiological Evidence for Supramodality in Human Metacognition. *The Journal of Neuroscience*, *38*(2), 263–277. https://doi.org/10.1523/JNEUROSCI.0322-17.2017

Faivre, N., Mudrik, L., Schwartz, N., & Koch, C. (2014). Multisensory integration in complete unawareness: evidence from audiovisual congruency priming. *Psychological Science*, *25*(11), 2006–2016. https://doi.org/10.1177/0956797614547916

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. Retrieved from http://www.jneurosci.org/content/22/13/5749

Fang, F., & He, S. (2005). Cortical responses to invisible objects in the human dorsal and ventral pathways. *Nature Neuroscience*, *8*(10), 1380–1385. https://doi.org/10.1038/nn1537

Filzmoser, P., Baumgartner, R., & Moser, E. (1999). A hierarchical clustering method for analyzing functional MR images. *Magnetic Resonance Imaging*, *17*(6), 817–826. https://doi.org/10.1016/S0730-725X(99)00014-4

Fleming, S. M., Huijgen, J., & Dolan, R. J. (2012). Prefrontal Contributions to Metacognition in Perceptual Decision Making. *Journal of Neuroscience*, *32*(18), 6117–6125. https://doi.org/10.1523/JNEUROSCI.6489-11.2012

Fort, A., Delpuech, C., Pernier, J., & Giard, M. H. (2002). Early auditory-visual interactions in human cortex during nonredundant target identification. *Cognitive Brain Research*, *14*(1), 20–30. https://doi.org/10.1016/S0926-6410(02)00058-7

Foxe, J. J., Wylie, G. R., Martinez, A., Schroeder, C. E., Javitt, D. C., Guilfoyle, D., … Murray, M.

M. (2002). Auditory-somatosensory multisensory processing in auditory association cortex: an fMRI study. *Journal of Neurophysiology*, *88*(1), 540–543. Retrieved from http://jn.physiology.org/content/88/1/540.long

Freeman, L. C. A., Wood, K. C., & Bizley, J. K. (2018). Multisensory stimuli improve relative localisation judgments compared to unisensory auditory or visual stimuli. *The Journal of the Acoustical Society of America*, *143*(6), EL516–EL522. https://doi.org/10.1121/1.5042759

Frens, M. A., & Van Opstal, A. J. (1998). Visual-auditory interactions modulate saccade-related activity in monkey superior colliculus. *Brain Research Bulletin*, *46*(3), 211–224. https://doi.org/10.1016/S0361-9230(98)00007-0

Frissen, I., Vroomen, J., de Gelder, B., & Bertelson, P. (2003). The aftereffects of ventriloquism: Are they sound-frequency specific? *Acta Psychologica*, *113*(3), 315–327. https://doi.org/10.1016/S0001-6918(03)00043-X

Galvin, S. J., Podd, J. V., Drga, V., & Whitmore, J. (2003). Type 2 tasks in the theory of signal detectability: Discrimination between correct and incorrect decisions. *Psychonomic Bulletin & Review*, *10*(4), 843–876. https://doi.org/10.3758/BF03196546

Gardner, W. G., & Martin, K. D. (1995). HRTF measurements of a KEMAR. *The Journal of the Acoustical Society of America*, *97*(6), 3907–3908. https://doi.org/10.1121/1.412407

Gau, R., & Noppeney, U. (2016). How prior expectations shape multisensory perception. *NeuroImage*, *124*, 876–886. https://doi.org/10.1016/j.neuroimage.2015.09.045

Gayet, S., Paffen, C. L. E., & Van der Stigchel, S. (2013). Information matching the content of visual working memory is prioritized for conscious access. *Psychological Science*, *24*(12), 2472–2480. https://doi.org/10.1177/0956797613495882

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. *The 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

Gherman, S., & Philiastides, M. G. (2015). Neural representations of confidence emerge from the process of decision formation during perceptual choices. *NeuroImage*, *106*, 134–143. https://doi.org/10.1016/j.neuroimage.2014.11.036

Giard, M. H., & Peronnet, F. (1999). Auditory-visual integration during multimodal object recognition in humans: a behavioral and electrophysiological study. *Journal of Cognitive Neuroscience*, *11*(5), 473–490. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/10511637

Grootswagers, T., Wardle, S. G., & Carlson, T. A. (2017). 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

Grossberg, S. (1999). The Link between Brain Learning, Attention, and Consciousness. *Consciousness and Cognition*, *8*(1), 1–44. https://doi.org/10.1006/ccog.1998.0372

Hackett, T. A., De La Mothe, L. A., Ulbert, I., Karmos, G., Smiley, J., & Schroeder, C. E. (2007). Multisensory convergence in auditory cortex, II. Thalamocortical connections of the caudal superior temporal plane. *The Journal of Comparative Neurology*, *502*(6), 924–952. https://doi.org/10.1002/cne.21326

Hairston, W. D., Laurienti, P. J., Mishra, G., Burdette, J. H., & Wallace, M. T. (2003). Multisensory enhancement of localization under conditions of induced myopia. *Experimental Brain Research*, *152*(3), 404–408. https://doi.org/10.1007/s00221-003-1646-7

Hairston, W. D., Wallace, M. T., Vaughan, J. W., Stein, B. E., Norris, J. L., & Schirillo, J. a. (2003). Visual Localization Ability Influences Cross-Modal Bias. *Journal of Cognitive Neuroscience*, *15*(1), 20–29. https://doi.org/10.1162/089892903321107792

Hammond-Kenny, A., Bajo, V. M., King, A. J., & Nodal, F. R. (2017). Behavioural benefits of multisensory processing in ferrets. *European Journal of Neuroscience*, *45*(2), 278–289. https://doi.org/10.1111/ejn.13440

Harris, J. A., Wu, C.-T., & Woldorff, M. G. (2011). Sandwich masking eliminates both visual awareness of faces and face-specific brain activity through a feedforward mechanism. *Journal of Vision*, *11*(7), 3–3. https://doi.org/10.1167/11.7.3

Hastie, T., Tibshirani, R., & Friedman, J. (2009). *The Elements of Statistical Learning*. *Revista Espanola de las Enfermedades del Aparato Digestivo* (Vol. 26). New York, NY: Springer New York. https://doi.org/10.1007/978-0-387-84858-7

Hautus, M. (2015). Signal Detection Theory. In *International Encyclopedia of the Social & Behavioral Sciences* (Second Edi, Vol. 21, pp. 946–951). Elsevier. https://doi.org/10.1016/B978-0-08-097086-8.43090-4

Helbig, H. B., & Ernst, M. O. (2007). Knowledge about a common source can promote visual-haptic integration. *Perception*, *36*(1972), 1523–1533. https://doi.org/10.1068/p5851

Heyman, T., & Moors, P. (2014). Frequent Words Do Not Break Continuous Flash Suppression Differently from Infrequent or Nonexistent Words: Implications for Semantic Processing of Words in the Absence of Awareness. *PLoS ONE*, *9*(8), e104719. https://doi.org/10.1371/journal.pone.0104719

Hsiao, J.-Y., Chen, Y.-C., Spence, C., & Yeh, S.-L. (2012). Assessing the effects of audiovisual semantic congruency on the perception of a bistable figure. *Consciousness and Cognition*, *21*(2), 775–787. https://doi.org/10.1016/j.concog.2012.02.001

Huckauf, A., & Heller, D. (2004). On the relations between crowding and visual masking. *Perception & Psychophysics*, *66*(4), 584–595. https://doi.org/10.3758/BF03194903

Hughson, W., & Westlake, H. D. (1944). Manual for program outline for rehabilitation of aural casualties both military and civilian. *Transactions of the American Academy of Ophthalmology & Otolaryngology*, *48*, (Suppl.), 1-15.

Izatt, G., Dubois, J., Faivre, N., & Koch, C. (2014). A direct comparison of unconscious face processing under masking and interocular suppression. *Frontiers in Psychology*, *5*(JUL), 1–11. https://doi.org/10.3389/fpsyg.2014.00659

Jack, C. E., & Thurlow, W. R. (1973). Effects of Degree of Visual Association and Angle of Displacement on the "Ventriloquism" Effect. *Perceptual and Motor Skills*, *37*(3), 967–979. https://doi.org/10.2466/pms.1973.37.3.967

Jackson, C. V. (1953). Visual factors in auditory localization. *Quarterly Journal of Experimental Psychology*, *5*(2), 52–65. https://doi.org/10.1080/17470215308416626

Jiang, Y., Costello, P., & He, S. (2007). Processing of Invisible Stimuli: Advantage of Upright Faces and Recognizable Words in Overcoming Interocular Suppression. *Psychological Science*, *18*(4), 349–355. https://doi.org/10.1111/j.1467-9280.2007.01902.x

Kaernbach, C. (1991). Simple adaptive testing with the weighted up-down method. *Perception & Psychophysics*, *49*(3), 227–229. https://doi.org/10.3758/BF03214307

Kanaya, S., & Yokosawa, K. (2011). Perceptual congruency of audio-visual speech affects ventriloquism with bilateral visual stimuli. *Psychonomic Bulletin & Review*, *18*, 123–128. https://doi.org/10.3758/s13423-010-0027-z

Kandil, F. I., Diederich, A., & Colonius, H. (2014). Parameter recovery for the time-window-of-integration (TWIN) model of multisensory integration in focused attention. *Journal of Vision*, *14*(11), 1–20. https://doi.org/10.1167/14.11.14

Kerkhoff, G. (2001). Spatial hemineglect in humans. *Progress in Neurobiology*, *63*(1), 1–27. https://doi.org/10.1016/S0301-0082(00)00028-9

Kim, C.-Y., & Blake, R. (2005). Psychophysical magic: rendering the visible 'invisible.' *Trends in Cognitive Sciences*, *9*(8), 381–388. https://doi.org/10.1016/j.tics.2005.06.012

King, J.-R., & Dehaene, S. (2014). Characterizing the dynamics of mental representations: The temporal generalization method. *Trends in Cognitive Sciences*, *18*(4), 203–210. https://doi.org/10.1016/j.tics.2014.01.002

King, J.-R., Pescetelli, N., & Dehaene, S. (2016). Brain Mechanisms Underlying the Brief Maintenance of Seen and Unseen Sensory Information. *Neuron*, *92*(5), 1122–1134. https://doi.org/10.1016/j.neuron.2016.10.051

Kingdom, F., & Prins, N. (2010). *Psychophysics: A Practical Introduction.* London: Academic Press: an imprint of Elsevier. Retrieved from http://www.palamedestoolbox.org

Kleiner, M., Brainard, D. H., & Pelli, D. G. (2007). What's new in Psychtoolbox-3? In *Perception, 36 (EVCP Abstract Supplement)* (p. Perception 36 ECVP Abstract Supplement). Alezzo.

Knotts, J. D., Lau, H., & Peters, M. A. K. (2018). Continuous flash suppression and monocular pattern masking impact subjective awareness similarly. *Attention, Perception, & Psychophysics*, *80*(8), 1974–1987. https://doi.org/10.3758/s13414-018-1578-8

Koivisto, M., & Rientamo, E. (2016). Unconscious vision spots the animal but not the dog: Masked priming of natural scenes. *Consciousness and Cognition*, *41*, 10–23.

https://doi.org/10.1016/j.concog.2016.01.008

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), 10. https://doi.org/10.1371/journal.pone.0000943

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

Lau, H. C., & Passingham, R. E. (2006). Relative blindsight in normal observers and the neural correlate of visual consciousness. *Proceedings of the National Academy of Sciences of the United States of America*, *103*(49), 18763–18768. https://doi.org/10.1073/pnas.0607716103

Laurienti, P. J., Kraft, R. a., Maldjian, J. a., Burdette, J. H., & Wallace, M. T. (2004). Semantic congruence is a critical factor in multisensory behavioral performance. *Experimental Brain Research*, *158*(4), 405–414. https://doi.org/10.1007/s00221-004-1913-2

Lee, A. L. F., Ruby, E., Giles, N., & Lau, H. (2018). Cross-Domain Association in Metacognitive Efficiency Depends on First-Order Task Types. *Frontiers in Psychology*, *9*(DEC), 1–10. https://doi.org/10.3389/fpsyg.2018.02464

Lee, H., & Noppeney, U. (2014). Temporal prediction errors in visual and auditory cortices. *Current Biology*, *24*(8), R309–R310. https://doi.org/10.1016/j.cub.2014.02.007

Lee, M., Blake, R., Kim, S., & Kim, C.-Y. (2015). Melodic sound enhances visual awareness of congruent musical notes, but only if you can read music. *Proceedings of the National Academy of Sciences*, *112*(22), 201509529. https://doi.org/10.1073/pnas.1509529112

Leek, M. R. (2001). Adaptive procedures in psychophysical research. *Perception & Psychophysics*, *63*(8), 1279–1292. https://doi.org/10.3758/BF03194543

Lemm, S., Blankertz, B., Dickhaus, T., & Müller, K. R. (2011). Introduction to machine learning for brain imaging. *NeuroImage*, *56*(2), 387–399. https://doi.org/10.1016/j.neuroimage.2010.11.004

Levitt, H. (1971). Transformed up-down methods in psychoacoustics. *The Journal of the Acoustical Society of America*, *49*(2), Suppl 2:467+. Retrieved from http://bdml.stanford.edu/twiki/pub/Haptics/DetectionThreshold/psychoacoustics.pdf

Lewald, J. (2002). Rapid Adaptation to Auditory-Visual Spatial Disparity. *Learning & Memory*, *9*(5), 268–278. https://doi.org/10.1101/lm.51402

Lewald, J., Ehrenstein, W. H., & Guski, R. (2001). Spatio-temporal constraints for auditory-visual integration. *Behavioural Brain Research*, *121*(1–2), 69–79. https://doi.org/10.1016/S0166-4328(00)00386-7

Lewald, J., & Guski, R. (2003). Cross-modal perceptual integration of spatially and temporally disparate auditory and visual stimuli. *Brain Research. Cognitive Brain Research*, *16*(3), 468–478. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/12706226

Lewald, J., & Guski, R. (2004). Auditory-visual temporal integration as a function of distance: No compensation for sound-transmission time in human perception. *Neuroscience Letters*, *357*(2), 119–122. https://doi.org/10.1016/j.neulet.2003.12.045

Lin, Y., Liu, B., Liu, Z., & Gao, X. (2015). EEG gamma-band activity during audiovisual speech comprehension in different noise environments. *Cognitive Neurodynamics*, *9*(4), 389–398. https://doi.org/10.1007/s11571-015-9333-5

Logothetis, N. K., Leopold, D. a, & Sheinberg, D. L. (1996). What is rivalling during binocular rivalry? *Nature*. https://doi.org/10.1038/380621a0

Lovelace, C. T., Stein, B. E., & Wallace, M. T. (2003). An irrelevant light enhances auditory detection in humans: a psychophysical analysis of multisensory integration in stimulus detection. *Cognitive Brain Research*, *17*(2), 447–453. https://doi.org/10.1016/S0926-6410(03)00160-5

Lu, L., Zhang, G., Xu, J., & Liu, B. (2018). Semantically Congruent Sounds Facilitate the Decoding of Degraded Images. *Neuroscience*, *377*, 12–25. https://doi.org/10.1016/j.neuroscience.2018.01.051

Ludwig, K., & Hesselmann, G. (2015). Weighing the evidence for a dorsal processing bias under continuous flash suppression. *Consciousness and Cognition*, *35*, 251–259. https://doi.org/10.1016/j.concog.2014.12.010

Ludwig, K., Kathmann, N., Sterzer, P., & Hesselmann, G. (2015). Investigating category- and shape-selective neural processing in ventral and dorsal visual stream under interocular suppression. *Human Brain Mapping*, *36*(1), 137–149. https://doi.org/10.1002/hbm.22618

Lunghi, C., & Alais, D. (2015). Congruent tactile stimulation reduces the strength of visual suppression during binocular rivalry. *Scientific Reports*, *5*, 9413. https://doi.org/10.1038/srep09413

Lunghi, C., Binda, P., & Morrone, M. C. (2010). Touch disambiguates rivalrous perception at early stages of visual analysis. *Current Biology*, *20*(4), R143–R144. https://doi.org/10.1016/j.cub.2009.12.015

Lunghi, C., Lo Verde, L., & Alais, D. (2017). Touch Accelerates Visual Awareness. *I-Perception*, *8*(1), 204166951668698. https://doi.org/10.1177/2041669516686986

Ma, W. J., Beck, J. M., Latham, P. E., & Pouget, A. (2006). Bayesian inference with probabilistic population codes. *Nature Neuroscience*, *9*(11), 1432–1438. https://doi.org/10.1038/nn1790

Macaluso, E., & Driver, J. (2005). Multisensory spatial interactions: A window onto functional integration in the human brain. *Trends in Neurosciences*, *28*(5), 264–271. https://doi.org/10.1016/j.tins.2005.03.008

Maniscalco, B., & Lau, H. (2012). A signal detection theoretic approach for estimating metacognitive sensitivity from confidence ratings - Supplementary Materials. *Consciousness and Cognition*.

Maniscalco, B., & Lau, H. (2014). Signal Detection Theory Analysis of Type 1 and Type 2 Data: Meta-d0, Response- Specific Meta-d0, and the Unequal Variance SDT Model. In Stephen M. Fleming & C. D. Frith (Eds.), *The Cognitive Neuroscience of Metacognition* (pp. 25–56). Berlin, Heidelberg: Springer. https://doi.org/10.1007/978-3-642-45190-4

Marsman, M., & Wagenmakers, E.-J. (2017). Bayesian benefits with JASP. *European Journal of Developmental Psychology*, *14*(5), 545–555. https://doi.org/10.1080/17405629.2016.1259614

Maruya, K., Watanabe, H., & Watanabe, M. (2008). Adaptation to invisible motion results in low-level but not high-level aftereffects. *Journal of Vision*, *8*(11), 7. https://doi.org/10.1167/8.11.7

McGurk, H., & MacDonald, J. (1976). Hearing lips and seeing voices. *Nature*, *264*(5588), 746–748. https://doi.org/10.1038/264746a0

Meijer, D., Veselič, S., Calafiore, C., & Noppeney, U. (2019). Integration of audiovisual spatial signals is not consistent with maximum likelihood estimation. *Cortex*, *119*, 74–88. https://doi.org/10.1016/j.cortex.2019.03.026

Mensen, A., & Khatami, R. (2013). Advanced EEG analysis using threshold-free cluster-enhancement and non-parametric statistics. *NeuroImage*, *67*, 111–118. https://doi.org/10.1016/j.neuroimage.2012.10.027

Meyerhoff, H. S., & Huff, M. (2015). Semantic congruency but not temporal synchrony enhances long-term memory performance for audio-visual scenes. *Memory & Cognition*, 390–402. https://doi.org/10.3758/s13421-015-0575-6

Molholm, S., Ritter, W., Javitt, D. C., & Foxe, J. J. (2004). Multisensory Visual-Auditory Object Recognition in Humans: A High-density Electrical Mapping Study. *Cerebral Cortex*, *14*(4), 452–465. https://doi.org/10.1093/cercor/bhh007

Moors, P., Boelens, D., van Overwalle, J., & Wagemans, J. (2016). Scene Integration Without Awareness. *Psychological Science*, *27*(7), 945–956. https://doi.org/10.1177/0956797616642525

Moors, P., Hesselmann, G., Wagemans, J., & van Ee, R. (2017). Continuous Flash Suppression: Stimulus Fractionation rather than Integration. *Trends in Cognitive Sciences*, *8*(0), 1380–1385. https://doi.org/10.1016/j.tics.2017.06.005

Morales, J., Lau, H., & Fleming, S. M. (2018). Domain-General and Domain-Specific Patterns of Activity Supporting Metacognition in Human Prefrontal Cortex. *The Journal of Neuroscience*, 2360–17. https://doi.org/10.1523/JNEUROSCI.2360-17.2018

Morein-Zamir, S., Soto-Faraco, S., & Kingstone, A. (2003). Auditory capture of vision: Examining temporal ventriloquism. *Cognitive Brain Research*, *17*(1), 154–163. https://doi.org/10.1016/S0926-6410(03)00089-2

Mudrik, L., Faivre, N., & Koch, C. (2014). Information integration without awareness. *Trends in Cognitive Sciences*, *18*(9), 488–496. https://doi.org/10.1016/j.tics.2014.04.009

Munhall, K. G., Gribble, P., Sacco, L., & Ward, M. (1996). Temporal constraints on the McGurk effect. *Perception & Psychophysics*, *58*(3), 351–362. https://doi.org/10.3758/BF03206811

Munhall, K. G., ten Hove, M. W., Brammer, M., & Paré, M. (2009). Audiovisual Integration of Speech in a Bistable Illusion. *Current Biology*, *19*(9), 735–739. https://doi.org/10.1016/j.cub.2009.03.019

Nahorna, O., Berthommier, F., & Schwartz, J.-L. (2012). Binding and unbinding the auditory and visual streams in the McGurk effect. *The Journal of the Acoustical Society of America*, *132*(2), 1061–1077. https://doi.org/10.1121/1.4728187

Nahorna, O., Berthommier, F., & Schwartz, J.-L. (2015). Audio-visual speech scene analysis: Characterization of the dynamics of unbinding and rebinding the McGurk effect. *The Journal of the Acoustical Society of America*, *137*(May), 362–377. https://doi.org/10.1121/1.4904536

Nelson, T. O. (1984). A comparison of current measures of the accuracy of feeling-of-knowing predictions. *Psychological Bulletin*, *95*(1), 109–133. https://doi.org/10.1037/0033-2909.95.1.109

Newman, J., & Baars, B. J. (1993). Aneural attentional model for access to consciousness: a global workspace perspective. *Concepts Neuroscience*, *4*, 255–290. Retrieved from http://cogprints.org/73/1/CINSART.htm

Ngo, M. K., & Spence, C. (2010). Crossmodal Facilitation of Masked Visual Target Discrimination by Informative Auditory Cuing. *Neuroscience Letters*, *479*(2), 102–106. https://doi.org/10.1016/j.neulet.2010.05.035

Ngo, M. K., & Spence, C. (2012). Facilitating masked visual target identification with auditory oddball stimuli. *Experimental Brain Research*, *221*(2), 129–136. https://doi.org/10.1007/s00221-012-3153-1

Nidiffer, A. R., Stevenson, R. A., Krueger Fister, J., Barnett, Z. P., & Wallace, M. T. (2016). Interactions between space and effectiveness in human multisensory performance. *Neuropsychologia*, *88*, 83–91. https://doi.org/10.1016/j.neuropsychologia.2016.01.031

Noguchi, Y., & Kakigi, R. (2005). Neural mechanisms of visual backward masking revealed by high temporal resolution imaging of human brain. *NeuroImage*, *27*(1), 178–187. https://doi.org/10.1016/j.neuroimage.2005.03.032

O'Toole, A. J., Jiang, F., Abdi, H., Pénard, N., Dunlop, J. P., & Parent, M. A. (2007). Theoretical, Statistical, and Practical Perspectives on Pattern-based Classification Approaches to the Analysis of Functional Neuroimaging Data. *Journal of Cognitive Neuroscience*, *19*(11), 1735–1752. https://doi.org/10.1162/jocn.2007.19.11.1735

Olivers, C. N. L., & Van der Burg, E. (2008). Bleeping you out of the blink: Sound saves vision from oblivion. *Brain Research*, *1242*, 191–199. https://doi.org/10.1016/j.brainres.2008.01.070

Oostenveld, R., Fries, P., Maris, E., & Schoffelen, J. M. (2011). FieldTrip: Open source software for advanced analysis of MEG, EEG, and invasive electrophysiological data. *Computational Intelligence and Neuroscience*, *2011*. 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*(July), 1–27. https://doi.org/10.3389/fninf.2016.00027

Palmer, T. D., & Ramsey, A. K. (2012). The function of consciousness in multisensory integration. *Cognition*, *125*(3), 353–364. https://doi.org/10.1016/j.cognition.2012.08.003

Pan, Y., Lin, B., Zhao, Y., & Soto, D. (2014). Working memory biasing of visual perception without awareness. *Attention, Perception, and Psychophysics*, *76*(7), 2051–2062. https://doi.org/10.3758/s13414-013-0566-2

Pápai, M. S., & Soto-Faraco, S. (2017). Sounds can boost the awareness of visual events through attention without cross-modal integration. *Scientific Reports*, *7*(January), 1–13. https://doi.org/10.1038/srep41684

Parise, C., & Spence, C. (2008). Synesthetic congruency modulates the temporal ventriloquism effect. *Neuroscience Letters*, *442*(3), 257–261. https://doi.org/10.1016/j.neulet.2008.07.010

Passamonti, C., Frissen, I., & Làdavas, E. (2009). Visual recalibration of auditory spatial perception: two separate neural circuits for perceptual learning. *European Journal of Neuroscience*, *30*(6), 1141–1150. https://doi.org/10.1111/j.1460-9568.2009.06910.x

Pereda, E., Quiroga, R. Q., & Bhattacharya, J. (2005). Nonlinear multivariate analysis of neurophysiological signals. *Progress in Neurobiology*, *77*(1–2), 1–37. https://doi.org/10.1016/j.pneurobio.2005.10.003

Pereira, F., & Botvinick, M. (2011). Information mapping with pattern classifiers: A comparative study. *NeuroImage*, *56*(2), 476–496. https://doi.org/10.1016/j.neuroimage.2010.05.026

Pereira, F., Mitchell, T., & Botvinick, M. (2009). Machine learning classifiers and fMRI: a tutorial overview. *NeuroImage*, *45*(1 Suppl), S199–S209. https://doi.org/10.1016/j.neuroimage.2008.11.007

Peremen, Z., & Lamy, D. (2014). Comparing unconscious processing during continuous flash suppression and meta-contrast masking just under the limen of consciousness. *Frontiers in Psychology*, *5*(AUG), 1–13. https://doi.org/10.3389/fpsyg.2014.00969

Pfurtscheller, G., Flotzinger, D., & Kalcher, J. (1993). Brain-Computer Interface—a new communication device for handicapped persons. *Journal of Microcomputer Applications*, *16*(3), 293–299. https://doi.org/10.1006/jmca.1993.1030

Radeau, M. (1974). Adaptation au déplacement prismatique sur la base d’une discordance entre la vision et l’audition. *L’ Année Psychologique*, *74*(1), 23–33. https://doi.org/10.3406/psy.1974.28021

Radeau, M., & Bertelson, P. (1977). Adaptation to auditory-visual discordance and ventriloquism in semirealistic situations. *Perception & Psychophysics*, *22*(2), 137–146. https://doi.org/10.3758/BF03198746

Radeau, M., & Bertelson, P. (1978). Cognitive factors and adaptation to auditory-visual discordance. *Perception & Psychophysics*, *23*(4), 341–343. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/748857

Ramsøy, T. Z., & Overgaard, M. (2004). Introspection and subliminal perception. *Phenomenology and the Cognitive Sciences*, *3*(1), 1–23. https://doi.org/10.1023/B:PHEN.0000041900.30172.e8

Raymond, J. E., Shapiro, K. L., & Arnell, K. M. (1992). Temporary suppression of visual processing in an RSVP task: An attentional blink? *Journal of Experimental Psychology: Human Perception and Performance*, *18*(3), 849–860. https://doi.org/10.1037/0096-1523.18.3.849

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. https://doi.org/10.1167/15.5.22

Rohe, T., & Noppeney, U. (2016). Distinct computational principles govern multisensory integration in primary sensory and association cortices. *Current Biology*, *26*(4), 509–514. https://doi.org/10.1016/j.cub.2015.12.056

Rohe, T., & Noppeney, U. (2018). Reliability-Weighted Integration of Audiovisual Signals Can Be Modulated by Top-down Control. *ENeuro*, *5*(1), ENEURO.0315-17.2018. https://doi.org/10.1523/ENEURO.0315-17.2018

Rothkirch, M., & Hesselmann, G. (2018). No evidence for dorsal-stream-based priming under continuous flash suppression. *Consciousness and Cognition*, *64*(May), 84–94. https://doi.org/10.1016/j.concog.2018.05.011

Rouault, M., McWilliams, A., Allen, M. G., & Fleming, S. M. (2018). Human Metacognition Across Domains: Insights from Individual Differences and Neuroimaging. *Personality Neuroscience*, *1*. https://doi.org/10.1017/pen.2018.16

Salimi-Khorshidi, G., Smith, S. M., & Nichols, T. E. (2011). Adjusting the effect of nonstationarity in cluster-based and TFCE inference. *NeuroImage*, *54*(3), 2006–2019. https://doi.org/10.1016/j.neuroimage.2010.09.088

Salomon, R., Galli, G., Łukowska, M., Faivre, N., Ruiz, J. B., & Blanke, O. (2016). An invisible touch: Body-related multisensory conflicts modulate visual consciousness. *Neuropsychologia*, *88*, 131–139. https://doi.org/10.1016/j.neuropsychologia.2015.10.034

Salomon, R., Kaliuzhna, M., Herbelin, B., & Blanke, O. (2015). Balancing awareness: Vestibular signals modulate visual consciousness in the absence of awareness. *Consciousness and*

Cognition, *36*, 289–297. https://doi.org/10.1016/j.concog.2015.07.009

Salomon, R., Lim, M., Herbelin, B., Hesselmann, G., & Blanke, O. (2013). Posing for awareness: Proprioception modulates access to visual consciousness in a continuous flash suppression task. *Journal of Vision*, *13*(7), 2–2. https://doi.org/10.1167/13.7.2

Salomon, R., Noel, J.-P. P., Łukowska, M., Faivre, N., Metzinger, T., Serino, A., & Blanke, O. (2017). Unconscious integration of multisensory bodily inputs in the peripersonal space shapes bodily self-consciousness. *Cognition*, *166*, 174–183. https://doi.org/10.1016/j.cognition.2017.05.028

Salti, M., Monto, S., Charles, L., King, J. R., Parkkonen, L., & Dehaene, S. (2015). Distinct cortical codes and temporal dynamics for conscious and unconscious percepts. *ELife*, *4*(MAY), 1–52. https://doi.org/10.7554/eLife.05652

Sandberg, K., Timmermans, B., Overgaard, M., & Cleeremans, A. (2010). Measuring consciousness: Is one measure better than the other? *Consciousness and Cognition*, *19*(4), 1069–1078. https://doi.org/10.1016/j.concog.2009.12.013

Sarmiento, B. R., Shore, D. I., Milliken, B., & Sanabria, D. (2012). Audiovisual interactions depend on context of congruency. *Attention, Perception, and Psychophysics*, *74*(3), 563–574. https://doi.org/10.3758/s13414-011-0249-9

Schmack, K., Burk, J., Haynes, J.-D., & Sterzer, P. (2016). Predicting Subjective Affective Salience from Cortical Responses to Invisible Object Stimuli. *Cerebral Cortex*, *26*(8), 3453–3460. https://doi.org/10.1093/cercor/bhv174

Schneider, T. R., Engel, A. K., & Debener, S. (2008). Multisensory identification of natural objects in a two-way crossmodal priming paradigm. *Experimental Psychology*, *55*(2), 121–132. https://doi.org/10.1027/1618-3169.55.2.121

Schroeder, C. E., & Foxe, J. J. (2002). The timing and laminar profile of converging inputs to multisensory areas of the macaque neocortex. *Cognitive Brain Research*, *14*(1), 187–198. https://doi.org/10.1016/S0926-6410(02)00073-3

Scott, R. B., Samaha, J., Chrisley, R., & Dienes, Z. (2018). Prevailing theories of consciousness are challenged by novel cross-modal associations acquired between subliminal stimuli. *Cognition*, *175*(February), 169–185. https://doi.org/10.1016/j.cognition.2018.02.008

Senkowski, D., Talsma, D., Grigutsch, M., Herrmann, C. S., & Woldorff, M. G. (2007). Good times for multisensory integration: Effects of the precision of temporal synchrony as revealed by gamma-band oscillations. *Neuropsychologia*, *45*(3), 561–571. https://doi.org/10.1016/j.neuropsychologia.2006.01.013

Senkowski, D., Talsma, D., Herrmann, C. S., & Woldorff, M. G. (2005). Multisensory processing and oscillatory gamma responses: effects of spatial selective attention. *Experimental Brain Research*, *166*(3–4), 411–426. https://doi.org/10.1007/s00221-005-2381-z

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). What you see is what you hear. *Nature*, *408*(6814), 788–788. https://doi.org/10.1038/35048669

Shams, L., Ma, W. J., & Beierholm, U. (2005). Sound-induced flash illusion as an optimal percept. *NeuroReport*, *16*(17), 1923–1927. https://doi.org/10.1097/01.wnr.0000187634.68504.bb

Shanks, D. R. (2017). Regressive research: The pitfalls of post hoc data selection in the study of unconscious mental processes. *Psychonomic Bulletin and Review*, *24*(3), 752–775. https://doi.org/10.3758/s13423-016-1170-y

Sheppard, J. P., Raposo, D., & Churchland, A. K. (2013). Dynamic weighting of multisensory stimuli shapes decision-making in rats and humans. *Journal of Vision*, *13*(6), 4. https://doi.org/10.1167/13.6.4

Sklar, A. Y., Levy, N., Goldstein, A., Mandel, R., Maril, A., & Hassin, R. R. (2012). Reading and doing arithmetic nonconsciously. *Proceedings of the National Academy of Sciences of the United States of America*, *109*(48), 19614–19619. https://doi.org/10.1073/pnas.1211645109

Skocik, M., Collins, J., Callahan-Flintoft, C., Bowman, H., & Wyble, B. (2016). I Tried a Bunch of Things: The Dangers of Unexpected Overfitting in Classification. *BioRxiv*, 078816. https://doi.org/10.1101/078816

Slutsky, D. a, & Recanzone, G. H. (2001). Temporal and spatial dependency of the ventriloquism effect. *Neuroreport*, *12*(1), 7–10. https://doi.org/10.1097/00001756-200101220-00009

Smith, S. M., & Nichols, T. E. (2009). Threshold-free cluster enhancement: Addressing problems of smoothing, threshold dependence and localisation in cluster inference. *NeuroImage*, *44*(1), 83–98. https://doi.org/10.1016/j.neuroimage.2008.03.061

Snoek, L., Miletić, S., & Scholte, H. S. (2019). How to control for confounds in decoding analyses of neuroimaging data. *NeuroImage*, *184*(September 2018), 741–760. https://doi.org/10.1016/j.neuroimage.2018.09.074

Stanislaw, H., & Todorov, N. (1999). Calculation of signal detection theory measures. *Behavior Research Methods, Instruments, & Computers*, *31*(1), 137–149. https://doi.org/10.3758/BF03207704

Stein, T., Hebart, M. N., & Sterzer, P. (2011). Breaking Continuous Flash Suppression: A New Measure of Unconscious Processing during Interocular Suppression? *Frontiers in Human Neuroscience*, *5*(December), 167. https://doi.org/10.3389/fnhum.2011.00167

Stein, T., & Sterzer, P. (2014). Unconscious processing under interocular suppression: Getting the right measure. *Frontiers in Psychology*, *5*(MAY), 1–5. https://doi.org/10.3389/fpsyg.2014.00387

Stein, T., Sterzer, P., & Peelen, M. V. (2012). Privileged detection of conspecifics: Evidence from inversion effects during continuous flash suppression. *Cognition*, *125*(1), 64–79. https://doi.org/10.1016/j.cognition.2012.06.005

Stekelenburg, J. J., Vroomen, J., & De Gelder, B. (2004). Illusory sound shifts induced by the ventriloquist illusion evoke the mismatch negativity. *Neuroscience Letters*, *357*(3), 163–166. https://doi.org/10.1016/j.neulet.2003.12.085

Stelzer, J., Chen, Y., & Turner, R. (2013). Statistical inference and multiple testing correction in classification-based multi-voxel pattern analysis (MVPA): Random permutations and cluster size control. *NeuroImage*, *65*, 69–82. https://doi.org/10.1016/j.neuroimage.2012.09.063

Sterzer, P., Haynes, J. D., & Rees, G. (2008). Fine-scale activity patterns in high-level visual areas encode the category of invisible objects. *Journal of Vision*, *8*(15), 10–10. https://doi.org/10.1167/8.15.10

Stevenson, R. A., & Wallace, M. T. (2013). Multisensory temporal integration: task and stimulus dependencies. *Experimental Brain Research*, *227*(2), 249–261. https://doi.org/10.1007/s00221-013-3507-3

Thagard, P., & Stewart, T. C. (2014). Two theories of consciousness: Semantic pointer competition vs. information integration. *Consciousness and Cognition*, *30*, 73–90. https://doi.org/10.1016/j.concog.2014.07.001

Thurlow, W. R., & Jack, C. E. (1973). Certain Determinants of the “Ventriloquism Effect.” *Perceptual and Motor Skills*, *36*(3_suppl), 1171–1184. https://doi.org/10.2466/pms.1973.36.3c.1171

Tononi, G., & Edelman, G. M. (1998). Consciousness and complexity. *Science*, *282*(5395), 1846–1851.

Tsilionis, E., & Vatakis, A. (2016). Multisensory binding: Is the contribution of synchrony and semantic congruency obligatory? *Current Opinion in Behavioral Sciences*, *8*, 7–13. https://doi.org/10.1016/j.cobeha.2016.01.002

Tsuchiya, N., & Koch, C. (2005). Continuous flash suppression reduces negative afterimages. *Nature Neuroscience*, *8*(8), 1096–1101. https://doi.org/10.1038/nn1500

Tsuchiya, N., Koch, C., Gilroy, L. a, & Blake, R. (2006). Depth of interocular suppression associated with continuous flash suppression, flash suppression, and binocular rivalry. *Journal of Vision*, *6*(10), 1068–1078. https://doi.org/10.1167/6.10.6

Vallar, G. (1998). Spatial hemineglect in humans. *Trends in Cognitive Sciences*, *2*(3), 87–97. https://doi.org/10.1016/S1364-6613(98)01145-0

Van Wanrooij, M. M., Bremen, P., & John Van Opstal, a. (2010). Acquired prior knowledge modulates audiovisual integration. *European Journal of Neuroscience*, *31*(10), 1763–1771. https://doi.org/10.1111/j.1460-9568.2010.07198.x

van Wassenhove, V., Grant, K. W., & Poeppel, D. (2007). Temporal window of integration in auditory-visual speech perception. *Neuropsychologia*, *45*(3), 598–607. https://doi.org/10.1016/j.neuropsychologia.2006.01.001

Vatakis, A., & Spence, C. (2008). Evaluating the influence of the “unity assumption” on the temporal perception of realistic audiovisual stimuli. *Acta Psychologica*, *127*(1), 12–23. https://doi.org/10.1016/j.actpsy.2006.12.002

Verhaal, J., & Luksch, H. (2016). Multimodal integration in the chicken. *The Journal of Experimental Biology*, *219*(Pt 1), 90–95. https://doi.org/10.1242/jeb.129387

Vidal, J. J. (1973). Toward Direct Brain-Computer Communication. *Annual Review of Biophysics and Bioengineering*, *2*(1), 157–180. https://doi.org/10.1146/annurev.bb.02.060173.001105

Vroomen, J., Bertelson, P., & de Gelder, B. (1998). A visual influence in the discrimination of auditory location. *Proceedings of the International Conference on Auditory-Visual Speech Processing AVSP'98*, 131–135. Retrieved from http://spitswww.uvt.nl/~vroomen/papers/proc 9.pdf

Vroomen, J., Bertelson, P., & De Gelder, B. (2001). The ventriloquist effect does not depend on the direction of automatic visual attention. *Perception & Psychophysics*, *63*(4), 651–659. https://doi.org/10.3758/BF03194427

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), 252–258. https://doi.org/10.1007/s00221-004-1899-9

Wang, Y., Celebrini, S., Trotter, Y., & Barone, P. (2008). Visuo-auditory interactions in the primary visual cortex of the behaving monkey: Electrophysiological evidence. *BMC Neuroscience*, *9*(1), 79. https://doi.org/10.1186/1471-2202-9-79

Weiskrantz, L. (1986). *Blindsight: A Case Study and Implications*. Oxford: Oxford University Press. Retrieved from http://www.oxfordscholarship.com/view/10.1093/acprof:oso/9780198521921.001.0001/acprof-9780198521921

Welch, R. B., & Warren, D. H. (1980). Immediate perceptual response to intersensory discrepancy. *Psychological Bulletin*, *88*(3), 638–667. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/7003641

Werner, S., & Noppeney, U. (2010). Distinct functional contributions of primary sensory and association areas to audiovisual integration in object categorization. *Journal of Neuroscience*, *30*(7), 2662–2675. https://doi.org/10.1523/JNEUROSCI.5091-09.2010

White, T. P., Wigton, R. L., Joyce, D. W., Bobin, T., Ferragamo, C., Wasim, N., … Shergill, S. S. (2014). Eluding the illusion? Schizophrenia, dopamine and the McGurk effect. *Frontiers in Human Neuroscience*, *8*(AUG), 1–12. https://doi.org/10.3389/fnhum.2014.00565

Wolpaw, J. R., Birbaumer, N., Heetderks, W. J., McFarland, D. J., Peckham, P. H., Schalk, G., … Vaughan, T. M. (2000). Brain-computer interface technology: a review of the first international meeting. *IEEE Transactions on Rehabilitation Engineering*, *8*(2), 164–173. https://doi.org/10.1109/TRE.2000.847807

Wolpaw, J. R., McFarland, D. J., Neat, G. W., & Forneris, C. A. (1991). An EEG-based brain-computer interface for cursor control. *Electroencephalography and Clinical*

Neurophysiology, *78*(3), 252–259. https://doi.org/10.1016/0013-4694(91)90040-B

Yuval-Greenberg, S., & Deouell, L. Y. (2007). What You See Is Not (Always) What You Hear: Induced Gamma Band Responses Reflect Cross-Modal Interactions in Familiar Object Recognition. *Journal of Neuroscience*, *27*(5), 1090–1096. https://doi.org/10.1523/JNEUROSCI.4828-06.2007

Yuval-Greenberg, S., & Heeger, D. J. (2013). Continuous Flash Suppression Modulates Cortical Activity in Early Visual Cortex. *Journal of Neuroscience*, *33*(23), 9635–9643. https://doi.org/10.1523/JNEUROSCI.4612-12.2013

Zakrzewski, A. C., Wisniewski, M. G., Iyer, N., & Simpson, B. D. (2019). Confidence tracks sensory- and decision-related ERP dynamics during auditory detection. *Brain and Cognition*, *129*(October 2018), 49–58. https://doi.org/10.1016/j.bandc.2018.10.007

Zhou, W., Jiang, Y., He, S., & Chen, D. (2010). Olfaction modulates visual perception in binocular rivalry. *Current Biology*, *20*(15), 1356–1358. https://doi.org/10.1016/j.cub.2010.05.059