MULTISENSORY INTEGRATION AND RECALIBRATION IN THE HUMAN BRAIN

by

DR MÁTÉ ALLER

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

November 2017

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This unpublished thesis/dissertation is copyright of the author and/or third parties.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

To cope with the challenges posed by our dynamically changing environment we rely on a number of senses as sources of information. The information provided by different senses must be seamlessly merged into an accurate and reliable percept at any moment throughout our lives regardless of the noisiness of our environment and the constantly changing nature of our sensory systems. Our understanding of these processes has expanded exponentially in recent decades; however there is an abundance of questions yet to be answered. The present thesis addresses some of the outstanding questions regarding multisensory integration and recalibration. In Chapter 1, we give an introduction to the background of multisensory integration. In Chapter 2 we review the neural mechanisms of auditory spatial perception. In Chapter 3 we lay methodological foundations for the empirical chapters. In Chapter 4, we investigate whether multisensory integration emerges prior to perceptual awareness. In Chapter 5, we scrutinize the neural dynamics of computations related to Bayesian Causal Inference. In Chapter 6 we examine the spatio-temporal characteristics of the neural processes of multisensory adaptation. Finally, in *Chapter* 7 we summarise the results of the empirical chapters, discuss their contribution to the literature and outline directions of future research.

For my mother

Édesanyámnak

ACKNOWLEDGEMENTS

I would like to thank, first and foremost, Uta Noppeney for her guidance and support throughout my PhD. Your way of thinking, the scientific rigour, foresight and creativity with which you approach research problems has had a great impact on me. I am continued to be stunned by the depth of your knowledge and I can only hope that someday – if I work hard – I will know as much as you do.

I would like to also thank my schoolmate, colleague and friend Ágoston Mihalik for sharing all the ups and downs of this journey with me. Without your help I probably would not have dared to pursue this PhD. I will miss the endless discussions sitting on the kitchen counter in Scott's Corner.

I would like to thank the present and previous members of the Computational Cognitive Neuroimaging Group (in no particular order): Arianna Zuanazzi, Patrycja Delong, David Meijer, Remi Gau, Steffen Burgers, Johanna Zumer, Tom White, Ambra Ferrari, Giulio Degano, Joana Leitao and Alex Krugliak for putting up with me and for creating a warm and friendly atmosphere both inside and outside the office. I have to particularly thank Sam Jones for proof reading and discussing my thesis.

Special thanks to all the members and staff of Second City Crossfit – my gym – for all the workouts we shared, which have helped keeping myself sane and healthy during these hard-working couple of years.

I would like to thank my mother Erika Bognár and all my family and friends back home in Hungary for the endless love, support and encouragement in pursuing my dream. Finally, I would like to say thank you to Tünde Endrei for her love and support and for being there all the time. I look forward to the future with you.

i

TABLE OF CONTENTS

Acknowledgementsi
List of figures vi
List of tables viii
Chapter 1: General introduction to multisensory integration1
How senses interact – an example2
Cognitive and stimulus-related factors in multisensory integration
The role of awareness in multisensory integration5
Optimality of multisensory integration7
Multisensory recalibration11
Neural substrates of multisensory processing14
Chapter 2: The neural mechanisms of auditory spatial perception
Cues for sound localization20
Neural pathways of auditory spatial processing22
Neural models of the representation of auditory space
Chapter 3: Methodological foundations
Multivairate decoding of neuroimaging data32
Representational similarity analysis
Bayesian Causal Inference43
Chapter 4: A spatially collocated sound thrusts a flash into awareness
Introduction

Methods54
Participants54
Stimuli and apparatus55
Experimental design57
Experimental procedure57
Analysis59
Results62
Effect of spatial congruency on visibility judgment62
Effect of sound location on perceived flash location as a function of visibility63
Discussion64
Chapter 5: To integrate or not to integrate: Temporal dynamics of Bayesian Causal
Inference
Introduction
Methods72
Participants72
Stimuli and apparatus73
Experimental design and procedure74
Eye movement recording and analysis75
EEG data acquisition76
General analysis strategy76
Behavioural analysis78

EEG data analysis79
Permutation of circular indices w_{AV} for behavioural and neural data81
Results
Behavioural results83
EEG results
Discussion94
Chapter 6: Spatio-temporal characteristics of visually induced auditory space
adaptation in the human brain97
Introduction97
Methods101
Behavioural experiment101
fMRI experiment111
EEG experiment118
Modelling124
Results125
Behavioural results128
Multivariate decoding results131
Characterization of neural representations of space and recalibration134
Discussion140
Chapter 7: General discussion146
Summary of main findings146

Chapter 2: A spatially collocated sound thrusts a flash into awareness146
Chapter 3: To integrate or not to integrate: Temporal dynamics of Bayesian
Causal Inference148
Chapter 4: Spatio-temporal characteristics of visually induced auditory space
adaptation in the human brain149
Contributions and future implications151
Conclusions155
References

LIST OF FIGURES

Chapter 1: General introduction

Figure 1.1 The maximum likelihood estimate

Chapter 2: The neural mechanisms of auditory spatial perception

Figure 2.1 Binaural cues for sound localization

Figure 2.2 The ascending auditory pathway in mammals

Figure 2.3 Coincidence detection according to the Jeffress model

Figure 2.4 The cross-correlogram

Figure 2.5 The population rate code model

Chapter 3: Methodological foundations

Figure 3.1 Data representation format for multivariate decoding

Figure 3.2 Representational dissimilarity matrix

Figure 3.3 The Bayesian Causal Inference model

Chapter 4: A spatially collocated sound thrusts a flash into awareness

Figure 4.1 Experiment paradigm and sample trial.

Figure 4.2 Behavioural results.

Chapter 5: To integrate or not to integrate: Temporal dynamics of Bayesian Causal Inference

Figure 5.1 Experimental design, example trial, and behavioural audiovisual weights.

Figure 5.2 Behavioural audiovisual weight index.

Figure 5.3 Behavioural responses and BCI model fits.

Figure 5.4 Time resolved decoding of stimulus location.

Figure 5.5 'Model free' EEG analysis results.

Figure 5.6 'Model based' EEG analysis results.

Chapter 6: Spatio-temporal characteristics of visually induced auditory space adaptation in the human brain

Figure 6.1 Overall layout, design, task and stimuli of the experiment.

Figure 6.2 Population rate code model.

Figure 6.3 Neuroimaging analysis pipeline

Figure 6.4 Behavioural results.

Figure 6.5 Multivariate decoding results.

Figure 6.6 Characterization of neural representations – fMRI.

Figure 6.7 Characterization of neural representations – EEG.

Figure 6.8 Second order MDS.

LIST OF TABLES

Chapter 4: Spatio-temporal characteristics of visually induced auditory space adaptation in the human brain

 Table 4.1 Behavioural results of the Bayesian model comparison between the null

 model and the full model.

Table 4.2 Behavioural results of hit rate, d' and PSE in the sound localization and adaptation tasks (group mean \pm SEM).

CHAPTER 1: GENERAL INTRODUCTION TO MULTISENSORY INTEGRATION

Human beings – as other living organisms – must survive in an ever-changing environment. As part of the animal kingdom we can actively interact with our surroundings in response to their present state. Information about the environment is therefore of utmost importance to us in choosing our actions, and ultimately in survival. Many million years of evolution equipped us with an array of senses through which we can gather information about the outside world. These senses are specialized to evaluate specific aspects of the environment and collectively they help to provide a unified, multisensory percept of the world around us. At first glance, the emergence of this multisensory percept seems trivial: our brains compute it effortlessly and continuously (at least during waking hours) throughout our lives. However, closer inspection reveals it to be anything but trivial. For example, what determines which pieces of information shall be combined from the deluge of input on a crowded crosswalk to avoid an approaching car? Or how does the brain differentiate the useful information from the useless noise? Questions like these have inspired many researchers over the past century to investigate the details of multisensory integration (Stein & Meredith, 1993). In this chapter I attempt to introduce the literature necessary to put into perspective the questions my colleagues and I investigated in the empirical chapters of this thesis: first we asked if a conscious signal in one modality can influence an unconscious signal in another modality (Chapter 4). Then, we investigated the temporal dynamics of Bayesian Causal Inference in the neocortex (Chapter 5). Finally we spatially and temporally

resolved the neural representation of auditory space and recalibration in the brain (Chapter 6).

How senses interact – an example

As an illustrative example let us consider the above scenario once again: we are crossing a busy street, some vehicles are parking, others are passing by. We would like to cross the street, so we look around and the way seems clear so we proceed to cross. Suddenly however we hear a car approaching from the right, which immediately catches our attention so we look to the right. We must decide whether we can cross the street safely before the car passes. Our decision crucially depends on what we see and hear. For example, if we hear and see the car slowing down and the turning signal flashing we can conclude that the car is taking a turn and we indeed have a clear way ahead. On the other hand, if the car is not slowing down and is instead headed towards us we must judge how fast is it approaching based on what we see and hear, allowing us to decide whether to step back or quickly proceed across. These two scenarios nicely illustrate the two general strategies our brain uses to combine information from different senses (Ernst & Bülthoff, 2004): one is to maximize the extracted information available in a given scenario (sensory *combination*), the other is to minimize the variability of some environmental property or – in other words – to increase its reliability (sensory integration). At first it was not clear if the car was headed towards us, so we used the increasing information from multiple sources – and combined these sources – to answer the question. This is an example of *disambiguation*. In the second scenario we were forced to judge how fast the car was approaching, so our brain used all the available information from both vision and hearing – integrating them – to come up with the most reliable estimate

possible of the car's speed. In the following sections we will discuss and explore properties of sensory integration in terms of these strategies.

Cognitive and stimulus-related factors in multisensory integration

The question of how senses interact was formally investigated as early as the end of the 19th century (Urbantschitsch, 1880) and has been of increasing interest ever since, particularly in recent decades. It is the general consensus amongst scientists that multisensory integration is dependent on both higher-order cognitive (top-down) and stimulus-related (bottom-up) factors (Gregg H. Recanzone, 2011; Talsma, Senkowski, Soto-Faraco, & Woldorff, 2010).

The role of attention in multisensory integration has been well established both at the behavioural and neuronal levels. It has been investigated in multiple experimental paradigms such as the McGurk effect (Alsius, Navarra, Campbell, & Soto-Faraco, 2005), the double flash and fusion illusions (Andersen, Tiippana, & Sams, 2004) and the ventriloquist effect (Busse, Roberts, Crist, Weissman, & Woldorff, 2005). It has been shown that both spatial (Fairhall & MacAluso, 2009; Santangelo, Belardinelli, Spence, & Macaluso, 2009; Santangelo & Macaluso, 2012) and modality-specific attention (Alsius, Navarra, & Soto-Faraco, 2007; Mozolic, Hugenschmidt, Peiffer, & Laurienti, 2008; Nardo, Santangelo, & Macaluso, 2014) have a modulatory effect. Other cognitive factors have also been investigated, such as the extent to which participants believe two signals have a common cause, also called the 'unity assumption' (Vatakis & Spence, 2008; Robert B. Welch, 1999). Furthermore, prior expectations (Van Wanrooij, Bremen, & John Van Opstal, 2010), emotional (Maiworm, Bellantoni, Spence, & Röder, 2012) and motivational (Bruns, Maiworm, & Röder, 2014) factors have also been shown to influence multisensory integration.

Finally we mention the role of awareness in multisensory integration (O. Deroy et al., 2016), which will be discussed in a later section.

Stimulus-related factors in multisensory integration were first investigated in anaesthetised animals using single-neuron recordings (Morrell, 1972; Stein, Magalhaes-Castro, & Kruger, 1975; Wickelgren, 1971). Three fundamental rules were identified from the individual neural response patterns: the spatial rule, the temporal rule, and the inverse effectiveness principle. The spatial rule states that integration is maximal when the various stimuli from different modalities emerge from the same spatial location, i.e. they activate neurons with overlapping receptive fields. The temporal rule states that integration is maximal when the individual sensory inputs arise approximately at the same time. Finally, according to the inverse effectiveness principle integration is maximal or most prominent when at least one of the sensory stimuli only elicits a weak response in the given neuron (Spence, 2013). Since these seminal studies the three rules have been examined in humans at the behavioural and neural levels. It has been shown that audiovisual integration falters as the spatial and/or temporal disparity between the individual stimuli increases (Gregg H. Recanzone, 2009; Slutsky & Recanzone, 2001; Wallace et al., 2004). However both the importance of the spatial rule (Spence, 2013) and the generalizability of the inverse effectiveness principle has been challenged recently (Holmes, 2007, 2009; Ma, Zhou, Ross, Foxe, & Parra, 2009; Stein, Stanford, Ramachandran, Perrault, & Rowland, 2009). Semantic congruency of audiovisual signals has been identified as a stimulus-related factor of integration as well (Calvert, Campbell, & Brammer, 2000; Krugliak & Noppeney, 2016; Laurienti, Kraft, Maldjian, Burdette, & Wallace, 2004; H. Lee & Noppeney, 2011; Van Atteveldt, Formisano,

Goebel, & Blomert, 2004); though it is important to note that evaluation of semantic congruency is dependent on higher order cognitive processes (Talsma et al., 2010).

The role of awareness in multisensory integration

It is an intriguing yet controversial question to what extent multisensory integration is automatic, i.e. dependent on awareness (O. Deroy et al., 2016; Ophelia Deroy, Chen, & Spence, 2014; Faivre, Mudrik, Schwartz, & Koch, 2014; Mudrik, Faivre, & Koch, 2014). Recent theories of consciousness suggest that global availability of information through long-range fronto-parietal connections is a prerequisite of information integration, which is considered as the basis of conscious experience (Baars, 2005; Dehaene & Naccache, 2001; Tononi & Edelman, 1998). It follows from this theory that conscious sensory information should be readily available to any other brain area specialized in the processing of other sensory modalities. Conversely, if sensory information does not reach consciousness it is confined in its respective primary sensory area. Based on this we would expect that conscious stimuli in one modality should influence the processing of an unconscious sensory input in another modality. In contrast, recent neural evidence suggests that crossmodal interactions emerge already at the primary sensory level (Foxe et al., 2002; A. A. Ghazanfar, 2005; Asif A. Ghazanfar & Schroeder, 2006; Kayser, Petkov, Augath, & Logothetis, 2005; Lakatos, Chen, O'Connell, Mills, & Schroeder, 2007; Hweeling Lee & Noppeney, 2014; Lehmann et al., 2006; Lewis & Noppeney, 2010; Rohe & Noppeney, 2015a, 2016; Schroeder & Foxe, 2002; Van Atteveldt et al., 2004) so it might also be possible that an unconscious sensory stimulus might influence another conscious or even an unconscious stimulus in a different modality. Indeed, in line with the first premise there is a large body of literature showing that aware stimuli

in one modality boost unaware stimuli in another modality into awareness; evidence span across different modality combinations and are in line with classical stimulusdriven mechanisms, such as spatial, temporal, semantic and phonological congruence (Adam & Noppeney, 2014; Alsius & Munhall, 2013; Y. C. Chen & Yeh, 2009; Y. C. Chen, Yeh, & Spence, 2011; V. Conrad, Bartels, Kleiner, & Noppeney, 2010; Verena Conrad, Vitello, & Noppeney, 2012; Verena Conrad et al., 2013; Hsiao, Chen, Spence, & Yeh, 2012; Lunghi & Alais, 2013; Lunghi, Binda, & Morrone, 2010; Lunghi, Morrone, & Alais, 2014; Lunghi, Verde, & Alais, 2017; Olivers & Van der Burg, 2008; Zhou, Jiang, He, & Chen, 2010). Most recently we demonstrated that a conscious auditory stimulus thrusts a suppressed visual stimulus into awareness depending on spatial congruency (Aller, Giani, Conrad, Watanabe, & Noppeney, 2015), a study which will be discussed in detail in *Chapter 4*. Evidence regarding the second prediction is much less firm. It has been shown that the McGurk illusion falters under continuous flash suppression (Palmer & Ramsey, 2012) and in a bistable illusion (Munhall, ten Hove, Brammer, & Paré, 2009). A recent study found congruency priming effects on conscious audioviual targets even if participants were unaware of the audiovisual primes (Faivre et al., 2014). Critically, however, the effect was only present when participants were previously trained. It must be noted that, while in these studies higher order semantic information was integrated, it is still unclear if low-level spatiotemporal information can be integrated outside of awareness. A seminal study (Bertelson, Pavani, Ladavas, Vroomen, & De Gelder, 2000) demonstrated in patients with spatial hemineglect that the ventriloguist effect prevailed on patients' neglected (unconscious) hemifield. This result must be interpreted with caution, however, as participants did not show a significant

ventriloquist effect in their intact hemifield. Finally, in a series of two experiments using the ventriloquist paradigm with continuous flash suppression my colleagues demonstrated that participants showed significant ventriloquist effect even if the visual stimuli were judged 'invisible' (Delong et al. in preparation). This effect persisted even in participants who were not better than chance when locating the flash when it was judged 'invisible'. To conclude, there is firm evidence that conscious stimuli can influence the processing of unconscious stimuli in another modality, but it is much less certain whether unconscious stimuli can modulate processing of other stimulus modalities and further theoretical and experimental research is needed to address this question.

Optimality of multisensory integration

We now turn our attention to precisely how the brain forms a coherent multimodal estimate about the world. Let us consider our real-world example of crossing the busy road once again: there is a car approaching us and we must estimate its speed to decide whether to go ahead or step back. We have both auditory and visual information about the car's speed. In an ideal (i.e. 'noiseless') world, these pieces of information – as they originate from the same object – are perfectly matched. However, in the real world this is more the exception than the rule: sensory estimates are constantly confounded by environmental and sensory noise, even if they originate from the same object. The brain must somehow resolve the conflict to make a decision and take an action. But how? Does it rely purely on the visual or the auditory percept, or use something in between? An early study investigated this problem in the context of judging the size of an object which participants concurrently saw and felt (Rock & Victor, 1964). The results showed a strong visual dominance

over the percept which was later termed 'visual capture'. Subsequent research led to the formulation of the 'modality appropriateness' hypothesis (R. B. Welch & Warren, 1986), which states that the more appropriate or precise modality in the task at hand is preferred in cross-sensory conflict. For example, in spatial localization vision is generally more precise than audition so vision is preferred (Bertelson & Aschersleben, 1998; Bertelson & Radeau, 1981), whereas in temporal judgement tasks it is the other way around (Shams, Kamitani, & Shimojo, 2000). The modality appropriateness hypothesis provides a rough qualitative solution to the problem, but does not give a quantitative explanation of how we are able to form a multisensory estimate which is maximally reliable, often more so than the individual unisensory estimates. *Reliability* is defined as the inverse variance of the probability distribution of the estimate from a given sensory modality, so increasing the reliability of a sensory signal is equivalent to decreasing its variance. The mathematically optimal way of calculating the integrated estimate with the lowest possible variance is called the Maximum Likelihood Estimate (MLE). An environmental property S can be represented by a sensory system as

(1)
$$\hat{S}_i = f_i(S)$$

where *f* is the transformation by which the system performs the estimation and *i* is the index of the sensory modality. Each estimate \hat{S}_i is confounded by noise. If the noises are independent and Gaussian with variance σ_i^2 and the Bayesian prior is uniform, then the MLE of the environmental property is

(2)
$$\hat{S} = \sum_{i} w_i \hat{S}_i \quad with \quad w_i = \frac{1/\sigma_i^2}{\sum_j 1/\sigma_j^2}$$

This is essentially the weighted sum of the individual – unimodal – estimates with weights proportional to their inverse variance (Figure 1.1) (Ernst & Bülthoff, 2004). In this case the variance of the final, multisensory (AB) estimate is given by

(3)
$$\sigma_{AB}^2 = \frac{\sigma_A^2 \sigma_B^2}{\sigma_A^2 + \sigma_B^2}$$

This variance is equal or lower than the individual sensory variances σ_A^2 and σ_B^2 . Ernst and Banks (2002) showed that humans integrate visual and haptic information consistent with the MLE in a visuo-haptic object size estimation task. Participants judged the size of an object by looking at it and touching it at the same time. Critically the reliability of the visual percept was manipulated by adding noise to it. Participants' percepts were dominated by the more reliable modality at any given time, i.e. the visual if there was no noise added to the display, and the haptic if the visual signal was sufficiently corrupted by noise. However, participants' size estimates were improved by the presence of both signals. Moreover, maximal gain in reliability occurred when the reliabilities of the two modalities were approximately matched, which is also predicted by the MLE model. Optimal integration according to the MLE principle has been demonstrated in other experimental paradigms as well, such as in an audiovisual spatial localization task (Alais & Burr, 2004) and a visuo-vestibular heading discrimination task (C. R. Fetsch, Turner, DeAngelis, & Angelaki, 2009).



Figure 1.1 The maximum likelihood estimate

The MLE model describes integration using only sensory (i.e. bottom-up) information. Previously we have summarized the body of evidence that multisensory integration depends on both bottom-up and top-down factors. To incorporate top-down factors into the model it must be extended to a Bayesian framework, where the sensory distributions (*likelihood*) are combined with the distribution describing prior knowledge about the environment (*prior*) to estimate the distribution of the stimuli (*posterior*) using Bayes' rule. It is important to point out that the Bayesian framework incorporates the MLE as a special case, i.e. if the prior distribution is uniform (also called the *flat prior*) the Bayesian estimate equals the MLE.

The Bayesian model computes the integrated estimate of the environmental property with minimal noise and uncertainty, given the sensory inputs and prior knowledge; however, there are situations when the integration of unisensory signals into a single estimate is disadvantageous (Ernst & Di Luca, 2012). In some situations, the two signals should be segregated, for example when there are two underlying causes; using our previous example: we want to make sure that we do not integrate the image of the car approaching us with the sound of another car parking behind us. The model must be further extended to incorporate uncertainty about the causal

structure of the environment. This problem was addressed by the introduction of a hierarchical *Bayesian Causal Inference* (BCI) model (Körding et al., 2007; Sato, Toyoizumi, & Aihara, 2007). In this model, the posterior probability of the causal structure is estimated given the signals and the *common source prior*. Then spatial estimates under fusion (commons source) and full segregation (independent sources) are weighted proportionally to the posterior probability of their underlying causal structure. This model has been further studied and extended on the behavioural level (Beierholm, Quartz, & Shams, 2009; Wozny, Beierholm, & Shams, 2008, 2010; Wozny & Shams, 2011a). Using functional magnetic resonance imaging (fMRI), our colleagues also demonstrated that the human brain performs BCI in a hierarchy of cortical regions in an audiovisual spatial localization task (Rohe & Noppeney, 2015a, 2016). Using the same experimental paradigm with electroencephalography (EEG), we examined how BCI estimates dynamically emerge in the human brain across time. *Chapter 5* discusses this experiment in detail.

Multisensory recalibration

So far, we have discussed how our brain integrates unisensory signals to resolve sensory conflict and form a coherent percept of the environment. Random discrepancies between sensory modalities are common; however they are typically small and temporary. These could be due to various environmental factors (shadows, reflections, lighting conditions, echoes etc.), or noise in the sensory system. These inconsistencies are overcome by sensory integration. Nonetheless *systematic* and *sustained* discrepancies between sensory signals are also common, e.g. wearing gloves, new glasses, a hearing aid etc. The most natural example is during

development: as we grow, our sensory organs grow and their relative positions change, which can affect various sensory estimates (e.g. interaural time difference). In these situations the brain *adapts* so that the conflicting sensory maps are brought back into correspondence (Ernst & Di Luca, 2012), a process called recalibration or in neuroscientific terms - *plasticity* (Held, 1965). In parallel with sensory integration, sensory recalibration has been investigated in various sensory modalities, e.g. visual (Adams, Banks, & Van Ee, 2001), auditory (M. Radeau & Bertelson, 1974; G. H. Recanzone, 1998), haptic (Burge, Girshick, & Banks, 2010) and vestibular (A. Zaidel, Turner, & Angelaki, 2011; Adam Zaidel, Ma, & Angelaki, 2013). As for the mechanisms for multisensory plasticity, initially a 'visual dominant adaptation' paradigm was proposed following the 'visual capture' hypothesis of multisensory integration (Rock & Victor, 1964). This was later challenged by evidence from visual recalibration (Atkins, Jacobs, & Knill, 2003; Lewald, 2002). Instead, following the idea of reliability-weighted cue integration, a 'reliability-based adaptation' paradigm was suggested (Burge et al., 2010; Witten & Knudsen, 2005). A study by Zaidel and colleagues (2011) investigated this question in a visuo-vestibular heading discrimination task. Human and non-human primates were exposed to systematically discrepant visual and vestibular heading cues whilst the reliability of the visual cues was manipulated and the perceptual adaptation of the individual cues was assessed. Importantly no feedback was provided about the accuracy of the stimuli, hence the adaptation was 'unsupervised'. Surprisingly, recalibration did not depend on the reliability of the visual cue, and was instead observed in both modalities with its magnitude following a fixed ratio. The authors hypothesised that the fixed ratio could express information about the relative accuracy of the sensory modalities (accuracy

is defined as the probability with which the signal represents the true magnitude of the physical property that it reflects (Ernst & Di Luca, 2012)). Crucially, the accuracy of the stimuli cannot be determined just from the sensory signals: some form of external feedback is required and such information can potentially be accumulated over an extended period. Unsupervised recalibration therefore acted as a way of reducing discrepancy between senses or maintaining the internal consistency of the system and depended on the relative accuracy of the modalities. The results are consistent with the evidence of visual dominant adaptation as well as it is a special case of fixed ratio adaptation (where the ratio goes to infinity for vision). Developmental studies have also confirmed the role of accuracy in recalibration (Gori, Del Viva, Sandini, & Burr, 2008; Gori, Sandini, Martinoli, & Burr, 2010). In a follow-up study Zaidel and colleagues (Adam Zaidel et al., 2013) investigated recalibration with external feedback (i.e. 'supervised recalibration') in the same experimental paradigm. Strikingly, their results showed that supervised recalibration depended on cue reliability: when the less reliable cue was inaccurate it alone got recalibrated. However, when the more reliable stimulus was inaccurate the two cues got calibrated together in the same direction. A hybrid model was proposed to reconcile the findings with previous evidence: supervised and unsupervised calibration act together simultaneously, with supervised calibration restoring the accuracy of the multisensory percept and unsupervised recalibration minimizing the discrepancy between the modalities.

Despite the extensive behavioural evidence, the neural mechanisms of multisensory recalibration remain unknown. To address some of the emerging questions we combined psychophysics, fMRI, EEG and multivariate pattern decoding to reveal the

neural mechanisms underlying the recalibration of auditory space after exposure to spatially discrepant audiovisual stimuli ('ventriloquist aftereffect'). *Chapter 6* discusses the study in detail.

Neural substrates of multisensory processing

Extensive research has been undertaken to explore the mechanisms of the interplay between senses, not only behaviourally but also at the neural level. Neuroscience has identified a number of brain areas where information from multiple senses converges (Felleman & Van Essen, 1991) and these areas were classically considered as the sites of multisensory integration. However recent research shed light on interactions between brain areas which were traditionally considered to be sensory-specific (Driver & Noesselt, 2008; Asif A. Ghazanfar & Schroeder, 2006). In this section I attempt to summarize the current views on the neural mechanisms of multisensory processing in general and of audiovisual processing in particular.

The prerequisite of multisensory processing is the convergence of inputs from distinct sensory modalities on individual neurons (Clemo, Keniston, & Meredith, 2011). Numerous anatomical and neurophysiological studies have demonstrated such connections in nonhuman primates and other mammals at subcortical and cortical levels. These have been mostly confirmed in humans using neuroimaging techniques (C Cappe, Rouiller, & Barone, 2012). Importantly, neuroimaging techniques such as fMRI or MEG/EEG are only capable of resolving population level activations in contrast to the single neuron level resolution of electrophysiology. This poses potential issues in differentiating multisensory convergence (i.e. different modalities project onto different cells within the population) from true multisensory integration (i.e. different modalities project onto the same neurons). Furthermore, modulatory

interactions are common both in higher-level (Avillac, Ben Hamed, & Duhamel, 2007; Barraclough, Xiao, Baker, Oram, & Perrett, 2005) and lower-level (Kayser, Petkov, & Logothetis, 2008; Meredith & Allman, 2009) regions. In these cases one modality (e.g. visual) only modulates the response elicited by another modality (e.g. auditory) and does not elicit a response on its own. Therefore care must be taken in designing and interpreting the results of these experiments (Noppeney, 2012).

On the subcortical level, the superior colliculus (SC) has been studied extensively in the context of sensory convergence. Seminal studies from Stein and colleagues have demonstrated that the cat SC, receives projections from auditory, visual and somatosensory areas (Meredith & Stein, 1983, 1986; Stein, 1978; Stein & Arigbede, 1972). Furthermore, the inverse effectiveness principle of multisensory integration also emerged from these influential studies (Stein & Meredith, 1993). Other subcortical regions such as the basal ganglia (Nagy, Eördegh, Paróczy, Márkus, & Benedek, 2006; von Saldern & Noppeney, 2013) and the thalamus (Céline Cappe, Morel, Barone, & Rouiller, 2009) have been shown to be involved in multisensory integration.

On the cortical level, the superior temporal sulcus (STS) has been shown to have multimodal neurons (Barraclough et al., 2005; Beauchamp, 2005; Bruce, Desimone, & Gross, 1981) and bidirectional connections with auditory, visual and somatosensory cortices (Padberg, Seltzer, & Cusick, 2003; Schmahmann & Pandya, 1991). The temporo-parietal junction (TPT) (Leinonen, Hyvarinen, & Sovijarvi, 1980) and areas of the intraparietal sulcus (IPS) are involved in multimodal space perception (Duhamel, Colby, & Goldberg, 1998; Rohe & Noppeney, 2015a, 2016). The premotor cortex (Graziano, Reiss, & Gross, 1999; Graziano, Yap, & Gross,

1994) and the ventrolateral prefrontal cortex (VLPFC) (Barbas et al., 2005; Romanski, 2007; Sugihara, Diltz, Averbeck, & Romanski, 2006) have been implicated in multisensory processing (Driver & Noesselt, 2008). These areas are considered the classical multisensory cortical areas and collectively are referred to as higher-order association cortex (Asif A. Ghazanfar & Schroeder, 2006). Together these results support the traditional model of multisensory integration where interaction between sensory information occurs only after extensive processing in the respective sensory-specific brain areas.

This view has been challenged more recently by a rapidly increasing body of evidence of intersensory interactions at the sensory-specific or even the primary sensory cortical levels. Although electrophysiological studies in the early 1970s already provided evidence of auditory responses in visual cortex (Fishman, Fishman, Michael, & Michael, 1973; Morrell, 1972; Spinelli, Starr, & Barrett, 1968), these were considered the result of confounding or nonspecific factors (Driver & Noesselt, 2008). Since then numerous studies have demonstrated such effects with state-of-the-art methods and design: invasive recordings in animals (Bizley, Nodal, Bajo, Nelken, & King, 2007; Brosch, 2005; Fu, 2004; A. A. Ghazanfar, 2005; Kayser et al., 2005; Lakatos et al., 2007; Werner-Reiss, Kelly, Trause, Underhill, & Groh, 2003), and fMRI (Amedi, 2002; Calvert, Hansen, Iversen, & Brammer, 2001; E. Macaluso, Frith, & Driver, 2000; Martuzzi et al., 2007; Watkins, Shams, Tanaka, Haynes, & Rees, 2006) and MEG/EEG (Burenko, 2010; Busse et al., 2005; Molholm et al., 2002; Murray et al., 2005; Teder-Sälejärvi, Di Russo, McDonald, & Hillyard, 2005; Teder-Sälejärvi, Di Russo, McDonald, & Hillyard, 2005; Teder-Sälejärvi, McDonald, Di Russo, & Hillyard, 2002) in humans.

Driver and Noesselt (Driver & Noesselt, 2008) summarized the emerging explanations in three possible accounts: the 'All multisensory' account proposed by (Asif A. Ghazanfar & Schroeder, 2006) is probably the most provocative, and states that all cortical regions are fundamentally multisensory. Although multisensory influences have been shown in all sensory specific cortices there is strong evidence that these areas show preference towards a particular modality (Emiliano Macaluso & Driver, 2005; Van Essen, Anderson, & Felleman, 1992). According to the 'New Bimodal Brain Areas' account, the newly found interactions would indicate the existence of not-yet-known convergence zones adjacent to the sensory specific areas (Beauchamp, Argall, Bodurka, Duyn, & Martin, 2004; Wallace et al., 2004). The third account stresses the 'Critical Role of Feedback Circuitry' and hypothesises that the interactions reflect feedback connections from traditional multisensory areas (Bonath et al., 2007; E. Macaluso et al., 2000; Emiliano Macaluso & Driver, 2005).

Numerous studies investigated the neural substrates of audiovisual integration specifically, with many focusing on the ventriloquist effect (Bischoff et al., 2007; Busse et al., 2005; Teder-Sälejärvi et al., 2005). The first study to demonstrate a link between the influence of the visual signal on auditory processing and the illusory percept was performed by Bonath and colleagues (2007). Using EEG and fMRI they showed that the illusory shift in sound location was associated with lateralized activity pattern in the auditory cortex of planum temporale (PT). EEG recordings revealed that the asymmetrical activity emerged between 230 – 270 ms post stimulus (Bonath et al., 2007), a finding which supports the role of cortical feedback mechanisms. A further study by the same group investigated the effect of temporal coincidence on the neural substrates of the ventriloquist illusion (Bonath et al., 2014). These results

confirmed that activations associated with the ventriloquist effect were observed in PT. Callan et al. (2015) recently have also investigated the spatial representation of auditory space and the ventriloquist effect using fMRI. They observed a monotonic increase in the blood oxygenation level-dependent (BOLD) signal for more contralateral sounds, which is consistent with the prevailing population rate code model of auditory space representation (McAlpine, 2005; Salminen, May, Alku, & Tiitinen, 2009; Stecker, Harrington, & Middlebrooks, 2005). Furthermore, the monotonic sensitivity of the BOLD signal to spatial location was attenuated by spatially discrepant visual stimuli (Callan et al., 2015). A very recent fMRI study confirmed that in nonhuman primates, auditory space is coded in a way consistent with the population rate code model (Ortiz-Rios et al., 2017).

In this chapter I reviewed the literature essential for the understanding of the subsequent empirical chapters. First, I introduced the problem of multisensory integration through an everyday illustrative example, which will be referred to in later chapters as well. Then I reviewed how high level cognitive and low level stimulus related factors influence multisensory perception. I discussed the role of awareness in multisensory processing before turning my attention to mathematical models describing optimal cue integration. Multisensory recalibration processes are closely related to integration, hence I devoted a section to review them. Finally I briefly summarized the main neural substrates underlying multisensory integration and recalibration.

CHAPTER 2: THE NEURAL MECHANISMS OF AUDITORY SPATIAL PERCEPTION

The experimental chapters of this thesis scrutinize the processes of multisensory integration in the context of auditory and visual space perception. The ability to localize the sources of sounds and visual events in space is fundamental for survival. The visual sensory epithelium, the retina, is organized such that adjacent receptors process information originating from adjacent spatial locations. Moreover, this topographic map of space is preserved in the neuroanatomical organization of the subsequent processing stages in the brain up to higher level visual areas (Wandell, Dumoulin, & Brewer, 2007). The encoding of space in the auditory system is much less straightforward and remains debated to this day. In hearing, the general organizing principle of the sensory epithelium is tonotopy: spatially adjacent receptors in the cochlea process sounds of adjacent frequency. This means that the brain needs to extract spatial cues from the entire spectrum of the sound to form an auditory representation of space. Although the tonotopic organization is prevalent in the neuroanatomy of the processing stages in the brain it does not exclude the possibility of a topographical representation of space as well. Indeed, early models of auditory space perception assumed a topographical place code in the brain (Jeffress, 1948). However more recent electrophysiological and neuroimaging evidence suggests a population rate code (Ortiz-Rios et al., 2017; Salminen et al., 2009; Trapeau & Schönwiesner, 2015; Werner-Reiss & Groh, 2008), where the group activity of two opponent populations of neurons, tuned for left and right side, encodes the spatial locations of sounds. No consensus has been reached yet on this matter.

In this chapter I summarize the physical and neuroanatomical bases of sound localization and review the various models of auditory space coding in the brain.

Cues for sound localization

Human listeners show remarkable sound localization performance and able to discriminate differences of just 1-2 degrees in the angular location in the horizontal plane (Mills, 1958). The auditory system critically relies on various physical cues to assess the location of sound sources. Let us consider a sound source located in the horizontal plane (elevation of 0°) at 45° to the right of the listener (azimuth of 45°) as depicted in Figure 2.1 As the sound approaches the listener it reaches the left ear slightly delayed (Δ t, Fiugre 2.1A) as the sound must travel slightly longer distance. This is called the interaural time difference (ITD); its value is maximal for sounds at ±90° azimuth and 0 for sounds at 0° or 180° azimuth. The head also dampens the travelling sound wave, hence the left ear will experience slightly smaller sound intensity than the right (Δ I, Fiugre 2.1B). This is referred to as interaural intensity difference (IID) or interaural level difference (ILD) if it is expressed in dB.



Figure 2.1 Binaural cues for sound localization: ITD (A) and ILD (B). Based on Grothe et al. (2010)

These are the two main cues the brain leverages to discriminate sound sources in the horizontal plane. These are also referred to as binaural cues, as signals from both ears are required to detect them. Furthermore, there are the so called spectral cues which are essentially distortions of the sound spectra upon interacting with the torso, head, ears and pinna. This filtering action can be characterized by taking the ratio of the spectra of the incoming sound measured at the source and upon reaching the eardrum. This is called head-related transfer function (HRTF). These cues can be detected by a single ear (i.e. monaural) and are essential to discriminate the elevation and front-back relation of the sound sources (Butler, 1969), but they contribute to localization in every direction as well (Batteau, 1967; Fisher & Freedman, 1968). In this chapter I am focusing mainly on auditory spatial processing in the horizontal plane, hence my discussion will focus on mechanisms involved in the processing of binaural spatial cues.

If we consider pure tones as sound sources several limitations arise from the perspective of binaural cues. One is that the wavelength of sounds below ~2kHz is comparable to the human head size. This means that sounds below this frequency can easily 'bypass' the head, subjected little to nothing to the shadowing effect. On the other hand the head size also puts a limit on the maximum perceivable ITD, which is around 690 μ s for a sound at \pm 90° azimuth. With sinusoidal sound sources the ITD is equivalent to a phase difference between the signals of the two ears (interaural phase difference, IPD). For example, for a 500 Hz pure tone, with a period of 2000 μ s, an ITD of 500 μ s is equivalent to 90° IPD (one quarter of the period). For lower frequencies this works well and provides an unambiguous cue for sound localization. However, for higher frequencies as the period of the tone decreases,

multiple full cycles can fit into the ITD window and the auditory system can no longer explicitly decide which cycle from form one ear matches a certain cycle from the other. This issue becomes relevant as the period of the sound becomes less than twice the maximum possible ITD, approximately 1380 μ s or equivalently its frequency is higher than around 725 Hz (Moore, 2013). To summarize, for pure tone localization, ILD is most useful at higher frequencies, whereas ITD is more effective at lower frequencies. This idea is called the 'duplex theory' of sound localization and was coined by Lord Rayleigh (1907) more than a century ago.

Neural pathways of auditory spatial processing

In this section I briefly describe the main processing stages of binaural spatial cues in the ascending auditory pathway in mammals and summarize the neural machineries involved in ILD and ITD processing. Figure 2.2 summarizes the mammalian ascending auditory pathway. Sounds in the environment elicit mechanical vibrations on the eardrum which spread through the chain of ossicles of the middle ear onto the fluid system of the cochlea of the inner ear. In the cochlea sensory hair cells transduce the mechanical vibrations into bioelectric signals, which are transmitted in the form of action potentials to the central nervous system via the auditory nerve. The first processing stage in the brain stem is the cochlear nucleus. The bushy cells (BC) of the ventral cochlear nucleus (VCN) are critical in the precise temporal processing of sound as their activity is phase-locked to the fine temporal structure of low frequency sounds (< 3kHz) or to the envelope of high frequency sounds (Grothe et al., 2010). The next synaptic stage is the superior olivary complex (SOC) where the detection of binaural cues takes place.



Figure 2.1 The ascending auditory pathway in mammals. For clarity, only the contralateral side is shown. Based on Grothe et al. (2010)

Specifically, the medial superior olive (MSO) and the lateral superior olive (LSO) are the main sites of the binaural comparisons underlying ILD and ITD processing respectively. The SOC project fibres to the dorsal nucleus of the lateral lemniscus (DNLL) and the inferior colliculus (IC) in the midbrain The IC is an important processing stage where several other auditory pathways converge. By this stage the detection and initial processing of the binaural cues has been completed, however further remodelling of the representational structure of space is possible depending on the task to which it will be used (Vonderschen & Wagner, 2014). The neural pathways ascend further and enter the thalamocortical system: the medial geniculate body (MGB) and finally reach the auditory cortex, where higher order functions such as localization attention control and formation of auditory objects takes place

(Chechik & Nelken, 2012; Itatani & Klump, 2011; Schindler, Herdener, & Bartels, 2013). Binaural cues contribute to these processes mainly through helping the segregation or binding other auditory cues. Besides the thalamocortical system, the IC also projects fibres to the superior colliculus (SC) which is a major site of information integration across the senses (Meredith & Stein, 1986).

Neural models of the representation of auditory space

As briefly mentioned in the introduction, the primary feature represented in the auditory system - from the cochlea up to at least the primary auditory cortex - is frequency. A great deal of our understanding of space perception comes from studies of vision and touch where space is directly represented in the sensory epithelium and this remains the main organizational principle in the dedicated processing stages in the brain. Based on these it is natural to assume such a topographical representation of auditory space might exist in the brain. Most of research in auditory localization focused on the perception of ITDs as the main cue for localization, hence my discussion will focus on the neural models of ITD processing. The first and foremost model to be mentioned here is the classic Jeffress model (Jeffress, 1948), which has been highly influential since its inception in the mid-20th century. The Jeffress model hypothesises a series of coincidence detectors in the mid-brain each receiving input from both ears. Each coincidence detector responds maximally to a specific time delay between the inputs from two ears. This is achieved by internally shifting the input of one ear with respect to the other – by means of different axonal conduction times - so that the external delay (i.e. ITD) is compensated for and the signals coincide on the detector (Figure 2.3). Different ITDs would then elicit maximal activation on different neurons in the array, creating a place map of ITD.


Figure 2.3 Coincidence detection according to the the Jeffress model. Based on Jeffress (1948)

To summarize the Jeffress model has three main components (Vonderschen & Wagner, 2014): (i) the coincidence detector, (ii) internal delay lines and (iii) an explicit map of ITD. This model inspired a whole family of models as we shall see. From a computational perspective the operation of coincidence detection is equivalent to cross-correlating the signals of the two ears, hence the family of these models of auditory space perception are also called cross-correlation models (Stern & Trahiotis, 1995). The more refined versions of the Jeffress model take into account the tonotopic anatomical arrangement of the neural pathway and assume a series of internal delay lines for each frequency channel, forming a two-dimensional representation of frequency and ITD, the cross-correlogram (Trahiotis, Bernstein, & Akeroyd, 2001). Several refinements have been introduced to the model to explain peculiar binaural psychophysical phenomena. For instance, a noise with a relatively long ITD of -1500 µs (i.e. leading on the left) and 500 Hz centre frequency is perceived on opposite sides by human listeners depending on the bandwidth. For relatively broad bandwidth (400 Hz) it is perceived on the correct left side, however for relatively narrow bandwidth (50 Hz) it is perceived on the (wrong) right side

(Thompson et al., 2006; Trahiotis & Stern, 1989; von Kriegstein, Griffiths, Thompson, & McAlpine, 2008). To explain this phenomenon two modifications have been introduced by the weighted image model by Stern et al. (1988). Firstly, at the level of ITD detection (MSO) extra weight is given to the ITDs which are more central ('centrality weighting'). This can be realized by having more coincidence detector units for shorter ITDs than for longer. Secondly, at the next level of ITD processing (IC) more weight is given to the ITDs which are consistent across frequencies ('straightness weighting'). Because of the periodic nature of the noise there is a secondary activation peak at +500 µs (see Figure 2.4). For a narrow band noise this secondary activation peak gets a stronger weight since it is closer to the centre, hence it will be perceived on the right side. For a broadband noise the consistent activation at -1500 µs across a broader range of frequency channels outweighs the secondary activation peak at +500 µs ('straightness weighting'), hence it will be lateralized to the (correct) left side.



Figure 2.4 Cross-correlogram. Shaded discs represent coincidence detectors across ITDs and frequency channels. The strength of the shading indicates the centrality weighting. Red dashed lines show the π -limit. Grey lines represent the evoked activations by a narrow band stimulus centred at 500 Hz. Based on von Kriegstein et al. (2008)

The Jeffress model gained its strongest support from electrophysiological studies in birds, specifically in barn owls. There is strong electrophysiological and anatomical evidence (Carr & Konishi, 1988, 1990; Reyes, Rubel, & Spain, 1996), that the avian nucleus laminaris (NL, the equivalent of MSO in mammals) represents auditory space in a place code with a series of delay lines. In mammals however, such strong evidence for the Jeffress model has been elusive. Most importantly, the topographic arrangement of ITD sensitive neurons had not been demonstrated convincingly (Grothe et al., 2010). Furthermore, in vivo recordings from a variety of mammalian species failed to demonstrate neurons tuned to ITDs beyond approximately half of the period of their respective characteristic frequency (Figure 2.4 dashed red lines) (Brand, Behrend, Marguardt, McAlpine, & Grothe, 2002; Hancock & Delgutte, 2004; Joris & Yin, 2007; McAlpine, Jiang, & Palmer, 2001). Incorporating this so-called πlimit into the model precludes the possibility of straightness weighting as simply there are no coincidence detectors on which the weighting could act (see figure). Correspondingly, the π -limit model predicts lateralization opposite to the weighted image model for broadband sounds with ITDs beyond the π -limit. This offers a testable hypothesis which has been investigated by Thompson et al (Thompson et al., 2006). and von Kriegstein et al. (2008) fMRI in human participants. Thompson et al. (Thompson et al., 2006) demonstrated, that the blood oxygenation leveldependent (BOLD) activation profile observed in the IC follows the predictions of the π -limit model: the peak of the activity switched from the contralateral to the ipsilateral

side relative to the leading ear depending on whether the magnitude of the ITDs was within or beyond the π -limit. von Kriegstein et al. (2008) used the same data to investigate this question in the auditory cortex. Similarly to the IC, ITDs within the π limit evoked greater cortical BOLD activation contralateral to the leading ear, however ITDs beyond the π -limit generated a balanced pattern of BOLD activity between the two hemispheres. Collectively these results demonstrated no evidence for straightness weighting at least regarding mechanisms contributing to the BOLD response and suggest, that the side on which a sound is perceived is not reflected in the lateralization of the cortical response. Another key observation from electrophysiological studies in mammals is that the ITD sensitive neurons tend to respond maximally to ITDs outside the physiological range based on the anatomy of the animal's head. This has been demonstrated in various mammalian species at multiple processing stages along the auditory pathway (Brand et al., 2002; Groh, Kelly, & Underhill, 2003; Middlebrooks & Knudsen, 1984; Tollin & Yin, 2002; Werner-Reiss & Groh, 2008; Woods, Lopez, Long, Rahman, & Recanzone, 2006). These results led to a re-evaluation of the viability of the place code for auditory space in mammals and a new model has been proposed (McAlpine et al., 2001). In this socalled population rate code model ITD is not represented by the peak response of sharply tuned neurons, instead it is coded in the changes in the spike rate of broadly tuned neurons. As the best ITD is placed outside the physiological range the sensitive slope of the response function falls on the range of relevant ITDs (Figure 2.5). It has also been observed, that the majority of neurons seem to be tuned to contralateral ITDs whereas the minority to ipsilateral (Woods et al., 2006). Thereby the average relative population activity of two channels (contralateral and ipsilateral)

encodes the horizontal location of sound. This idea has been first proposed by von Békésy (1930).



Figure 2.5 Population rate code model. Grey lines represent the tuning curves of individual neurons. Based on Salminen et al. (2009)

Accumulating evidence from electrophysiological studies from macaque monkeys also support the population rate code model both at the level of the midbrain (Groh et al., 2003) and auditory cortex (Werner-Reiss & Groh, 2008). Deciphering these neural mechanisms in humans proved to be challenging as the available noninvasive methods are only capable of resolving summed responses of large populations of neurons. Salminen et al (2009). addressed the question whether space is coded in place or in rate code in the human auditory cortex using magnetoencephalography (MEG). In a stimulus-specific adaptation paradigm they investigated the effect of adaptors at different spatial locations to the N1 response (a prominent peak in the auditory event related potential at around 100 ms post stimulus) evoked by a probe. Based on previous evidence it is expected, that the attenuation of the N1 response is maximal (Butler, 1972) when the adaptor is at the

same location as the probe as both are stimulating the same neural population. The attenuation then decreases as the spatial disparity increases as different neural populations get activated by the adaptor and the probe. Critically, the place code and the population rate code have different predictions with respect to the profile of the attenuation. The place code predicts, that the magnitude of the attenuation progressively increases with spatial disparity but does not depend on the absolute location of the two stimuli. The rate code however predicts that the attenuation profile is mainly determined by whether the two stimuli are in the same or opposite hemifields. For stimuli in the same hemifield the neural populations are largely overlapping, hence the attenuation is high. Conversely, sound sources in the opposite hemifield activate different neural populations, therefore the attenuation is minimal. Using a variety of adaptor locations, the authors compared the attenuation profile of the N1 response to the predictions based on a place code and the rate code. The results supported the population rate code. Furthermore, modelling analyses suggested, that the contralaterally tuned population was larger than the ipsilaterally tuned.

Recent studies using fMRI in monkeys and humans also corroborated that auditory space is coded using a population rate code in the neocortex (Ortiz-Rios et al., 2017; Trapeau & Schönwiesner, 2015). In summary these findings are in contrast with the sparse topographical representation of auditory space which has its roots in the Jeffress model. To synthetize these pieces of contradicting evidence Vonderschen and Wagner (2014) proposed, that a distinction has to be made in the neural coding at the detection and the later remodelling of ITDs. The remodelling can change the code substantially along the ascending auditory pathway and there is evidence in

multiple species, including humans, that the representation of ITDs is remodelled into a few broadly tuned ITD channels irrespective of the representation at the detection stage (Harper & McAlpine, 2004). Collectively, more recent evidence supports the population rate code as the model of auditory space representation over the topographical place code in the human and primate neocortex. We choose this model to in conjunction with representational similarity analysis to characterize the representation of auditory space and recalibration in the human neocortex in chapter

6.

CHAPTER 3: METHODOLOGICAL FOUNDATIONS

In this chapter I discuss the non-standard methodologies used in the subsequent experimental chapters. In the first section I introduce the foundations and rationale of multivariate decoding analysis in general and the application of support vector machines, (SVM) particularly support vector regression (SVR) to the analysis of neuroimaging data (EEG and fMRI). In the second section I discuss representational similarity analysis (RSA) and how it can be applied to characterize the structure of the representations of phenomena in the brain and computational models, and quantitatively relate them across measured imaging modalities and models. In the final section I discuss Bayesian Causal Inference.

Multivairate decoding of neuroimaging data

The classical approach to the analysis of functional imaging data for many years has been *mass-univariate* methods. In this framework, one aims to understand how neural activity varies when there is concurrent variation in the world (Naselaris, Kay, Nishimoto, & Gallant, 2011), in other words, how the variation of the world (external variables) is *encoded* in the measured activity. Another characteristic of this approach is that each of the measurement channels (i.e. voxels in fMRI or electrodes in EEG) are related individually to a specific external variable (Haufe et al., 2014), essentially ignoring the rich information present in the spatio-temproal pattern of the data. In contrast, a different and more recently widespread approach is to look at the data and determine how much can be learned about the world (Naselaris et al., 2011), i.e. how much of the variability of the world (external variables) can be *decoded* from the brain activity. This approach combines activity across the measurement channels and utilizes the multivariate pattern present in the data to

approximate or decode the external variables. The usage of multivariate decoding or multivariate pattern analysis (MVPA) has been increased dramatically over the past decade and it is becoming a standard tool for functional imaging.

Let us consider a simple flash localization experiment as an illustrative example where participants locate flashes presented at different locations along the horizontal plane and their brain activity is recorded with fMRI or EEG. The type of questions we can answer using MVPA can be summarized as follows (Pereira, Mitchell, & Botvinick, 2009): First, is there any information regarding the variable of interest in the brain, i.e. can we decode the spatial location of the flash (pattern discrimination)? If yes, then we can ask where (e.g. which brain region in case of fMRI) or when (e.g. at what post stimulus time in case of EEG) this information can be decoded from the brain (pattern localization). Finally, we can examine the format in which the spatial location of the flash is represented (pattern characterization). My discussion in the current section covers the first two questions, while the third question will be covered in the section devoted to representational similarity analysis.

At the core of multivariate decoding there is the decoder, a function which maps the multidimensional data patterns to a variable of interest. At the technical level multivariate decoding belongs to the broader set of machine learning techniques. In the machine learning terminology, the multivariate data patterns are called features and the corresponding values of the variable of interest are the labels. An independent set of features is called an example (see Figure 3.1). In our illustrative example the features are fMRI voxel patterns or EEG topographies, the labels are the spatial locations of the flashes and the examples are the individual trials or individual runs (for details see later). If the labels are on a continuous scale (i.e. in

our example spatial location), the decoding is a regression problem, if the labels are discrete values (e.g. left vs. right), then it is a classification problem.



Figure 3.1 Data representation format for multivariate decoding

Briefly, the decoding process has the following steps: The decoder is first trained to learn the mapping from features to labels using examples with known labels (training set). This is called supervised learning as the decoder is informed about the correct labels. (There are also non-supervised and semi-supervised learning methods, which I am not discussing here as it would exceed the scope of the current chapter.) Then the learned mapping can be used to predict the labels of examples which were previously not seen by the decoder (test set) and the performance can be assessed by comparing the predicted labels to the real labels and an accuracy measure can be computed (typically percentage of correct predictions in case of classification or mean squared error in case of regression). Finally, the achieved accuracy can be tested using nonparametric permutation or bootstrap tests to see if the variable of interest could be successfully decoded (i.e. accuracy better than chance level).

After giving a summary of the decoding process I am now discussing each step of the process in more detail loosely following the discussion of Pereira et al. (2009)

- 1. The first step is transforming the data into examples (example creation). In general, examples can be created from single trials, blocks or runs. The choice must be carefully considered based on the imaging modality, experimental design and research question. In fMRI for example, the raw BOLD activations can be used for decoding in blocked designs, whereas in rapid event-related designs beta images after classical GLM estimation are preferred to account for the temporal autocorrelation due to the sluggish nature of the BOLD response. In EEG usually single trials or averaged potentials (e.g. over runs) are used. There is a trade-off between having many noisy examples or fewer cleaner ones (Grootswagers, Wardle, & Carlson, 2017). The most important consideration here is that examples must be independent of each other, otherwise the decoding gives overly optimistic estimates. Furthermore, each experimental condition must have equal number of examples otherwise the learning algorithm may get biased towards to most numerous condition (Pereira et al., 2009). These factors must be considered in the experimental design, making sure that conditions are equally likely to occur, and proper counterbalancing must be applied whenever necessary.
- 2. The next step is feature selection and processing. Neuroimaging modalities generate excessive amounts of data and usually there are much more features (i.e. voxels) than examples which renders the computational problem 'ill posed'. To mitigate this, it can be beneficial to reduce the number of features. In fMRI, one way is to apply some sort of dimensionality reduction like PCA (Hansen et al., 1999), ICA (Calhoun & Adali, 2006). Alternatively, the analysis can be restricted to a specific region of interest (ROI). It is common

practice to select voxels which show some activity compared to baseline during the task. Most importantly, it is essential that our selection criterion must not depend on the decoding labels, i.e. it is not acceptable to select those voxels, which appear to discriminate between conditions or classes. Once again, this would lead to over-optimistic training accuracies and poor generalization accuracy. In EEG, when time resolved decoding is applied the features typically comprise the topographies at a selected post stimulus time point or time window (Grootswagers et al., 2017). Since the number of EEG channels are typically less than 256 it is rarely necessary to further reduce the number of features. It is also highly recommended, that all features should be brought into a common scale. Typically features are normalized to have a mean of 0 and standard deviation of 1 either within each example (rows of the data matrix, see Figure 3.1), or across examples (columns of the data matrix, Figure 3.1). The former method is effective in reducing image- or topographywide variations, while the latter can be useful if the signal amplitude is more variable across time at some voxels/channels than the others.

3. Choosing the decoder. There is a plethora of decoding models to choose from. In general, decoders can be classified into discriminative and generative models (Hastie, Tibshirani, & Friedman, 2009). Discriminative models such as logistic regression or support vector machines (SVMs) learn the parameters of a prediction function directly from the training data so the parameters are set such that the prediction function predicts the desired label. Generative models, such as linear discriminant analysis (LDA) on the other hand learn a statistical model that can generate examples with the desired label. Although the

performance of nonlinear decoders can exceed those of linear, in neuroimaging research linear decoders are generally preferred over nonlinear ones mainly because the interpretation of the relationship between features and predictions is much more complicated in the latter case. Linear SVMs are very popular choice in neuroimaging research and have been implemented in numerous MVPA toolboxes for neuroimaging Pronto (Schrouff et al., 2013), TDT (Hebart, Görgen, & Haynes, 2015) and MNE (Gramfort et al., 2014) and I will discuss them in more details below.

4. The next step is the training and testing of the decoder. As mentioned above the training and testing must be done on distinct set of examples. In classical machine learning the examples are divided into 3 distinct sets: training, testing and validation. The training and testing sets are used to fine tune the hyperparameters of the decoder, so its performance is optimized for the data set at hand. The validation set is left untouched during the training-testing process and used only once to assess the performance of the final model. Good performance both on the testing and validation sets means that the decoder is appropriately fine tuned to the dataset at hand, furthermore it can generalize to unseen examples. Conversely, good training but poor validation performance means that the decoder is fine tuned to the training set to such an extent, that it cannot longer generalize to unseen examples, a phenomenon called overfitting. In neuroimaging research data are usually expensive to obtain, hence it would be uneconomical to leave data out from the analysis just for validation only. Furthermore, generally we are interested in the average performance of the decoder on the whole dataset, not just on a

subset of it. To address this, a process called cross-validation (CV) is generally used. In CV, the examples are randomly divided into several sets (e.g. 5) of equal size, called folds. On each iteration of the CV one decoder is trained such that one fold is selected for testing and the rest of the folds are used for training. This is then repeated systematically choosing a different fold for testing at every iteration. Finally, the average decoding performance is assessed by averaging the decoders' performance across folds. This is called the k-fold CV, where k stands for the number of folds. The means by which the left out examples at every CV iteration is selected trials, a single sample or a single run (keeping the trials of the run together is essential in fMRI decoding because of temporal autocorrelations) can be left out, these are called leave-one-sample-out (LOSO) and leave-one-run-out (LORO) respectively. When decoding from fMRI data with event related design, LORO-CV is generally preferred whereas with EEG generally k-fold CV is preferred.

5. Statistical inference: To assess if the performance of the decoder is better than chance, generally non-parametric permutation and bootstrap tests are preferred over traditional parametric t-tests or binomial tests as they have minimal assumptions with respect to the underlying distribution of the data. (Allefeld, Görgen, & Haynes, 2016; Haynes, 2015; Nichols & Holmes, 2001; Pereira & Botvinick, 2011; Stelzer, Chen, & Turner, 2013). The critical step in permutation tests is how the null distribution is generated, against which the observed decoding accuracy will be compared. On the single subject level inference, one can permute the labels, essentially disrupting the association

between labels and examples. This is then repeated many times with newly permuted labels yielding the null distribution. On the second, betweensubjects level, a sign permutation test can be used, where the subject-level decoding accuracies are randomly multiplied by +1 or -1 and a statistic of interest is computed based on the permuted samples. Repeating this process multiple times generates the null distribution.

This brief overview is hopefully successfully highlights the advantages and potential pitfalls of multivariate decoding analyses and at the same time it is useful for the understanding of Chapter 5 and 6, where these approaches formed the backbone of our analyses.

Representational similarity analysis

In the previous section I briefly touched upon temporal and across condition generalization which are examples of pattern characterization since they allow for the assessment of the temporal stability of patterns and how they change across conditions. These methods however are specific to imaging modalities with rich spatiotemporal resolution (e.g. EEG or MEG) and not allow for the direct examination of representational structure in the underlying activity patterns. In recent years a method called representational similarity analysis (RSA) has been developed by Kriegeskorte and colleagues (Kriegeskorte, 2008; Kriegeskorte & Kievit, 2013; Nili et al., 2014) which allows the characterization of representational structure in a wild variety of data ranging from behavioural, neuroimaging, electrophysiological and computational modelling and across species.

The core concept of RSA is the representational dissimilarity matrix (RDM). Treating the activation patterns related to a given experimental condition as a representation (i.e. manifestation of the mental state induced by the experimental condition) we can compare these in a pairwise fashion across all experimental conditions. This yields a comprehensive characterization of the representations across the set of experimental conditions. Each cell of the RDM represents the dissimilarity (or distance) between the respective pair of conditions. The RDM is symmetric about the diagonal, which consist of zeros (as it shows the distance between identical conditions, Figure 3.2). The RDM is an intuitive visual image and analysis tool of the representational structure, which – importantly – is free from the activation pattern-specific details, hence it opens up possibilities to compare representations across a wild array of data types (Kriegeskorte, 2008). Specifically, it allows (i) comparison of computational model predictions to brain activity, (ii) contrasting representations across brain regions, species and imaging modalities, (iii) relating brain and behaviour.



Figure 3.2 Representational dissimilarity matrix

Below I summarize the typical workflow of RSA based on Kriegeskorte (2008), using one of our analyses detailed in Chapter 6 as an illustrative example. Briefly, human participants localized sounds appearing from 7 locations along the horizontal plane and brain activations were recorded with fMRI and EEG in separate sessions.

- Estimation of activity patterns: Depending on the species and data recording technique the data could range from invasive electrophysiological recordings to neuroimaging and any other brain activity measurement technique. In our experiment condition-specific (i.e. across the 7 spatial locations) fMRI response patterns were estimated using the general linear model (GLM) for each region of interest (ROI), and condition-specific evoked potentials were computed from the epoched EEG responses.
- 2. Measuring dissimilarity in activation patterns: The dissimilarity between each pair of conditions-specific activity patterns is computed next. Any distance measure can be used to express the dissimilarity: often the Euclidean-, Mahalanobis- or correlation-distance (1 Pearson R) is used. In our experiment we use the Mahalanobis-distance. The dissimilarities were assembled into an RDM, separately for each ROI (fMRI) and time point (EEG).
- 3. Predicting representational similarity structure from computational model: Computational models of various complexities can be used here. We used the population rate code model (Salminen et al., 2009) to predict the activation patterns of two hypothetical populations of neurons tuned to the ipsi- and contralateral hemifields respectively. The predicted single neuron activations extracted at the 7 spatial locations formed the condition-specific activation patterns which were used to compute the model's RDM.
- 4. Comparing brain and model RDMs: For quantitative comparison the RDMs can also be considered as 'activation patterns' and the pairwise distance

between them can be computed the same way as between the 'real' activation patterns. Since the RDMs are symmetric, usually the lower triangular part is used as a pattern vector. Then the distances can be assembled into a secondorder RDM where each cell represents the distance between a pair of firstorder RDMs. Alternatively, a simple Spearman correlation can be computed between the first order RDMs which expresses the similarity between them. In our analysis we used both methods: we computed the distances between the representational structures observed in the brain regions (fMRI) and the population rate code model. Furthermore, we assessed how representations observed with fMRI corresponded to those observed with EEG by computing Spearman's correlation coefficient separately for each ROI between the RDM of the respective ROI and the RDM from each time point in the EEG time course.

5. Testing the relatedness of RDMs: To statistically test if two RDMs are similar (related) one can use an elegant randomization approach. First, the similarity is computed between the two RDMs using Spearman's correlation. Then one of the RDMs is permuted, meaning that the association between the condition labels and the dissimilarity between them is disrupted. By repeating this process many times (e.g. 10000) and computing the correlation with the permuted RDM yields a null distribution of correlation coefficients under the null hypothesis, that they are not related. Then the observed correlation between the original RDMs can be compared to the null-distribution and the null hypothesis is rejected at level α if the observed correlation falls within the top $\alpha \times 100\%$ of the null-distribution.

6. Visualization: RDMs convey information about the representational structure in an intuitive and visually pleasing manner. An alternative way of visualization is through a process called multidimensional scaling (MDS) (Shepard, 1980). This method essentially projects the distances from the original high dimensional similarity space to a lower dimension (typically 2D). Similar entities are clustered close, while dissimilar ones lie farther away.

This step-by-step guide introduces RSA and lays the foundations for the understanding of the analyses applied in Chapter 6. However it must be noted, that we only covered a subset of the potential analysis possibilities using RSA. For a more comprehensive review of all the possibilities of this method we refer to the following papers (Kriegeskorte, 2008; Kriegeskorte & Kievit, 2013; Nili et al., 2014).

Bayesian Causal Inference

Let us consider our illustrative example from Chapter 1 once again with a slight twist: we are about to cross a busy road in dusk and the road is covered in thick fog (a typical Birmingham late-afternoon in the winter). The position of an approaching car can only be guessed by its silhouette, lights and sound. The brain needs to accurately and reliably estimate the car's position at every moment to get us through the road safely, utilizing every piece of useful sensory information: the hazy image and sound. Two fundamental computational problems need to be solved: First, it must be determined whether these sensory inputs originate from the same car or different independent ones (i.e. common or sources). Spatio-temporal correspondences between the sensory stimuli as well as prior knowledge about the

environment inform the brain whether signals are likely to come from a common source. Second, the image and the sound should be integrated or segregated depending on the most likely causal structure: if they are caused by the same car (common source) the brain can use these redundant pieces of information to come up with the most reliable estimate of the car's position. Critically however, signals from different cars (independent sources) should be segregated (Figure 3.3).



Figure 3.3 The Bayesian Causal Inference model. Auditory (X_A) and visual (X_V) sensory representations emerge from a single (S_{AV}) or multiple (S_V, S_A) underlying environmental sources, depending on the causal structure (C).

Hierarchical Bayesian Causal Inference (BCI) (Körding et al., 2007; Wozny et al., 2010) provides a unifying framework to incorporate both sensory integration and segregation in perception: first it models the two possible causal structures (i.e. common or independent) that could give rise to the sensory inputs. In line with In case of a common source, signals are integrated weighted by their relative sensory reliabilities. In case of independent sources, they are segregated and processed separately. Critically, the brain has only access to noisy sensory information about the environment, including its causal structure. To account for the uncertainty, the estimates (e.g. the object's location) under the two causal structures are weighted in

the final estimate by their causal structure's posterior probability (for other decisional strategies see Wozny et al. (2010)).

The generative model of Bayesian Causal Inference (Figure 3.3) assumes that common (C = 1) or independent (C = 2) causes are sampled from a binomial distribution defined by the common cause prior P_{common} . For a common source, the 'true' location S_{AV} is drawn from the spatial prior distribution N(μ_{AV} , σ_P). For two independent causes, the 'true' auditory (S_A) and visual (S_V) locations are drawn independently from this spatial prior distribution. For the spatial prior distribution, usually a central bias is assumed (i.e., $\mu = 0$) as it is generally more likely for a stimulus to appear centrally than far to the side. Sensory noise is introduced by drawing x_A and x_V independently from normal distributions centred on the true auditory (resp. visual) locations with parameters σ_A^2 (resp. σ_V^2). Thus, the generative model includes the following free parameters: the common source prior p_{common}, the spatial prior variance σ_P^2 , the auditory variance σ_A^2 and the visual variance σ_V^2 . The posterior probability of the underlying causal structure can be inferred by combining the common-source prior with the sensory evidence according to Bayes rule:

(1)
$$p(C = 1|x_{A}, x_{V}) = \frac{p(x_{A}, x_{V}|C=1)p_{common}}{p(x_{A}, x_{V})}$$

In the case of a common source (C = 1), the optimal estimate of the audiovisual location is a reliability-weighted average of the auditory and visual percepts and the spatial prior.

(2)
$$\hat{S}_{AV,C=1} = \frac{\frac{X_A}{\sigma_A^2} + \frac{X_V}{\sigma_V^2} + \frac{\mu_P}{\sigma_P^2}}{\frac{1}{\sigma_A^2} + \frac{1}{\sigma_V^2} + \frac{1}{\sigma_P^2}}$$

In the case of independent sources (C = 2), the optimal estimates of the auditory and visual signal locations (for the auditory and visual location report, respectively) are independent from each other.

(3)
$$\hat{S}_{A,C=2} = \frac{\frac{X_A}{\sigma_A^2} + \frac{\mu_P}{\sigma_P^2}}{\frac{1}{\sigma_A^2} + \frac{1}{\sigma_P^2}}, \quad \hat{S}_{V,C=2} = \frac{\frac{X_V}{\sigma_V^2} + \frac{\mu_P}{\sigma_P^2}}{\frac{1}{\sigma_V^2} + \frac{1}{\sigma_P^2}}$$

To provide a final estimate of the auditory and visual locations, the brain can combine the estimates under the two causal structures using various decision functions such as 'model averaging', 'model selection' and 'probability matching' (Wozny et al., 2010). In Chapter 5, we present results using 'model averaging'. According to the 'model averaging' strategy, the brain combines the integrated forced fusion spatial estimate with the segregated, task-relevant unisensory (i.e., either auditory or visual) spatial estimates weighted in proportion to the posterior probability of the underlying causal structures.

(4)
$$\hat{S}_A = p(C=1|x_A, x_V) \hat{S}_{AV,C=1} + (1 - p(C=1|x_A, x_V)) \hat{S}_{A,C=2}$$

(5)
$$\hat{S}_V = p(C=1|x_A, x_V) \hat{S}_{AV,C=1} + (1 - p(C=1|x_A, x_V)) \hat{S}_{V,C=2}$$

To arbitrate between those three models (i.e. fusion, segregation, BCI), each model is fitted individually to a participant's behavioural localization responses (or neural response decoded from neuroimaging data, see Chapter 5 for EEG and Rohe and Noppeney (2015a) for fMRI) based on the predicted distributions of the auditory spatial estimates (i.e., $p(\hat{S}_A|S_A,S_V)$) and the visual spatial estimates (i.e., $p(\hat{S}_V|S_A,S_V)$). These distributions are generated by simulating x_A and x_V many times (e.e. 10000) for each of the experimental conditions and inferring \hat{S}_A and \hat{S}_V from equations (1)-(5), then marginalizing over the internal variables x_A and x_V that are not accessible to the experimenter.

(6)
$$p(\hat{S}_{V}|S_{A},S_{V}) = \iint p(\hat{S}_{V}|x_{V},x_{A})p(x_{V}|S_{V})p(x_{A}|S_{A})dx_{V}dx_{A}$$

To link $p(\hat{S}_A|S_A,S_V)$ and $p(\hat{S}_V|S_A,S_V)$ to discrete auditory and visual discrete responses at the behavioural level, it isassumed that participants selected the button that is closest to \hat{S}_A or \hat{S}_V , binning the \hat{S}_A and \hat{S}_V accordingly into a histogram (with bins corresponding to the response buttons). Thus, a histogram of predicted auditory or visual localization responses for each condition and participant is obtained. Based on these histograms the probability of a participant's counts of localization responses is computed using the multinomial distribution for a given condition:

(7)
$$p(\{n_i\}|\{p_i\}) = \frac{n!}{\prod_{i=1}^m n_i!} \prod_{i=1}^m p_i^{n_i}$$

where n is to total number of trials in this condition and m is the number of response options. This gives the likelihood of the model given participants' response data.

(8)
$$L(model|\{n_i\}) \propto \prod_{i=1}^m p_i^{n_i}$$

Assuming independence of conditions, the log likelihoods are summed across conditions. To obtain maximum likelihood estimates for the parameters of the models (p_{common} , σ_P , σ_A , σ_V ; formally, the forced fusion and segregation models assume

 $p_{common} = 1$ or = 0, respectively), a non-linear simplex optimization algorithm can be used, for example as implemented in Matlab's fminsearch function (Matlab R2016a).

To identify the optimal model for explaining participants' data, the candidate models are compared using e.g. the Bayesian information criterion (BIC) as an approximation to the model evidence (Kass & Raftery, 1995). The BIC depends on both model complexity and model fit. In Chapter 5 we performed Bayesian model selection (Rigoux, Stephan, Friston, & Daunizeau, 2014) at the group level as implemented in SPM8 (Friston, Holmes, et al., 1994) to obtain the protected exceedance probability for the candidate models.

CHAPTER 4: A SPATIALLY COLLOCATED SOUND THRUSTS A FLASH INTO AWARENESS

The material of this chapter is the result of a collaboration between Máté Aller, Anette Giani, Verena Conrad, Masataka Watanabe, Beatrix Barth, Natalie Christner and Uta Noppeney, and has been published in Aller et al. (2015) . The experiment was designed by AG, VC and UN, data was collected by BB and NC. All data analysis was performed by MA (supervised by UN), the manuscript was written by MA and UN.

Introduction

For effective interactions an organism needs to merge signals from different senses into a coherent and unified percept of the environment. A controversial question is to which extent multisensory integration is automatic or relies on higher cognitive resources such as attention or awareness (for review see Talsma, Senkowski, Soto-Faraco, & Woldorff, 2010). Even though recent studies have demonstrated that awareness and attention can be dissociated (Koch & Tsuchiya, 2007, 2012; M. Watanabe et al., 2011; Wyart & Tallon-Baudry, 2008), in many situations attention and awareness are closely intertwined. Hence, for the purpose of this study we do not yet intend to dissociate these aspects, but loosely define 'automatic integration' as integration that is relatively immune to attention and awareness. According to the account of automatic integration multisensory co-stimulation increases the bottom-up stimulus saliency (Onat, Libertus, & König, 2007). Thus, signals that co-occur within a spatial and temporal window of integration can automatically amplify stimulus

salience. Multisensory integration thereby enables multisensory events to enter perceptual awareness and capture an organism's attention.

In support of automatic integration a vast body of psychophysics and neurophysiological research has shown that multisensory integration is immune to attentional modulation (Bresciani, Dammeier, & Ernst, 2006; Stekelenburg, Vroomen, & De Gelder, 2004; Vroomen, Bertelson, & De Gelder, 2001), emerges prior to participants' awareness (Alsius & Munhall, 2013) and even persists in the anaesthetized non-human primate brain (e.g. superior colliculus, primary sensory areas) (Kayser et al., 2005; Stanford, 2005). Yet, the account of 'attention-free' integration has more recently been challenged. For instance, the audiovisual McGurk illusion falters, when attention is diverted to a secondary task (Alsius et al., 2005) or when subjects are unaware of the visual speech gestures (Munhall et al., 2009). Moreover, neuroimaging studies have shown profound attentional modulation of neural multisensory integration indices. Thus, attention modulated the amplification of the BOLD response for congruent audiovisual speech signals in superior colliculi, primary sensory and association cortices (Fairhall & MacAluso, 2009). Likewise, EEG studies showed attentional influences on audiovisual interactions already at \leq 100 ms poststimulus (Talsma, Doty, & Woldorff, 2007). With respect to perceptual awareness, the role of primary sensory areas is still debated. While numerous studies have demonstrated that activations in primary sensory areas correlate with participants' awareness (Tong, 2003), others have suggested that these activations may be mediated by concurrent attentional effects (M. Watanabe et al., 2011). Collectively, this body of research suggests a multifaceted and not yet completely

understood interplay between multisensory integration and higher cognitive processes such as attention or awareness (Talsma et al., 2010).

This intricate relationship partly results from the hierarchical nature of multisensory perception where different types of information (e.g. temporal, spatial, semantic, phonological) are integrated at distinct cortical levels (Bonath et al., 2007; Driver & Noesselt, 2008; H. Lee & Noppeney, 2011; Hweeling Lee & Noppeney, 2014; Lewis & Noppeney, 2010; Werner & Noppeney, 2010). Conversely, perceptual awareness and attentional capture rely on a cascade of neural processes. Thus, experiments using masking (Y. C. Chen & Spence, 2011), attentional blink (Adam & Noppeney, 2014; Olivers & Van der Burg, 2008; Soto-Faraco & Spence, 2002), binocular/perceptual rivalry (Alais, van Boxtel, Parker, & van Ee, 2010; V. Conrad et al., 2010; Verena Conrad et al., 2013, 2012; Guzman-Martinez, Ortega, Grabowecky, Mossbridge, & Suzuki, 2012; Hupé, Joffo, & Pressnitzer, 2008; Klink, van Wezel, & van Ee, 2012; Lunghi & Alais, 2013; Lunghi et al., 2010, 2014; Morrone & Lunghi, 2013; van Ee, van Boxtel, Parker, & Alais, 2009; Zhou et al., 2010) or flash suppression (Alsius & Munhall, 2013; Palmer & Ramsey, 2012) are likely to perturb the interplay between perceptual awareness and multisensory integration at different processing stages (for related discussion focusing on visual context, see Fogelson, Kohler, Miller, Granger, & Tse, 2014; Peremen & Lamy, 2014; for a recent review see Deroy et al., 2014). In particular, using binocular rivalry numerous studies have demonstrated that a concurrent non-visual signal increases the dominance and decreases the suppression times of the congruent visual percept. Yet, because of the presence of two rivaling percepts, these binocular rivalry experiments make it more difficult to unambiguously determine that the rivalry dynamics was shaped by

interactions between the non-visual signals with the suppressed rather than the dominant percept (for further discussion, please see Conrad et al., 2010).

Continuous flash suppression (CFS) is a powerful technique to manipulate participants' perceptual awareness (Tsuchiya & Koch, 2005). Flashing a mask to one eye can render even a salient stimulus presented to the other eye invisible. Critically, CFS is thought to affect cortical activity already at the primary cortical level via a gain control mechanism (Yuval-Greenberg & Heeger, 2013). CFS thus provides a very useful paradigm to investigate whether a concurrent non-visual signal can counteract the effect of flash suppression at the primary cortical level. Indeed, a previous study has demonstrated that an auditory speech signal makes participants more likely to detect a congruent relative to an incongruent speech video under continuous flash suppression (Alsius and Munhall, 2013; see also Palmer and Ramsey, 2012). These results suggest that audiovisual synchrony and temporal correlations are important determinants for audiovisual interactions prior to participants' awareness. Moreover, as natural speech signals evolve continuously over time, temporal expectations may also play an important role in enabling participants to detect visual speech signals.

Yet, as this previous study has presented auditory and visual signals only in a spatially congruent fashion, it could not evaluate the role of spatial congruency, which is another critical cue for multisensory binding. Spatial congruency may enable multisensory interactions via at least two mechanisms. First, spatial congruency may act as a bottom-up cue informing the brain that two signals are likely to come from a common source and should hence be bound into a coherent percept. Second, a spatially collocated sound may reduce observers' uncertainty about where a concurrent flash was presented. Even though spatial congruency affects detection

performance only rarely in redundant target paradigm (Bertini, Leo, & Làdavas, 2008; Forster, Cavina-Pratesi, Aglioti, & Berlucchi, 2002) the second mechanism may be more important in paradigms where the visual signal has been strongly attenuated by various experimental manipulations such as flash suppression or masking. Spatial uncertainty may be reduced via bottom-up mechanisms that enable the formation of more precise audiovisual spatial salience maps. Alternatively, a co-located sound may reduce spatial uncertainty even via top-down expectations that stabilize visual representations potentially even after they have accessed awareness.

Previous studies have demonstrated that a sound increases the detectability of a collocated yet masked visual flash at threshold visibility (Bolognini, Frassinetti, Serino, & Làdavas, 2005; Frassinetti, Bolognini, & Làdavas, 2002). Yet, as these masking studies reduced flash detectability only to threshold performance of 70%, the suppression of awareness for the undetected stimuli was rather shallow. Moreover, it is still unknown whether masking and dynamic continuous flash suppression reduce visual awareness via similar neural mechanisms (Fogelson et al., 2014; Peremen & Lamy, 2014).

To further investigate the role of spatial congruency in multisensory integration prior to perceptual awareness, the current study combined spatial audiovisual stimulation with dynamic continuous flash suppression (Maruya, Watanabe, & Watanabe, 2008; Tsuchiya & Koch, 2005). On each trial, participants were presented with a single flash in the centre, their left or right hemifield together with a sound that was spatially congruent or incongruent. Participants located the flash (i.e. flash localization) and judged its visibility (i.e. visual detection task). First, we investigated whether participants were better at detecting the flash when the sound was spatially

collocated. We hypothesized that spatial constraints are critical for audiovisual integration processes prior to participants' awareness. Second, we investigated whether the concurrent sound biased participants' perceived flash location and whether this bias depended on flash visibility. Importantly, as CFS obliterated visual awareness only in a fraction of trials, we were able to compare the audiovisual spatial bias for physically identical flashes that were visible or invisible.



Figure 4.1 Experiment paradigm and sample trial. (A) Experiment design 2 x 3 factorial design with factors: (i) Sound location: left, right; (ii) Flash location: left, centre, right. (B) Example trial and procedure of dynamic flash suppression

Methods

Participants

After giving informed consent, 24 healthy young adults with normal or corrected-tonormal vision participated in this study (14 females, mean age: 26.7 years, standard deviation: 5.3, range: 18-40; 22 right-handed). One subject was excluded because she did not follow task instructions properly as she located the visual stimuli almost exclusively in the centre (98.5%, (group mean \pm SD): 35.7% \pm 17.5%). The study was approved by the local ethics review board of the University of Tübingen.

Stimuli and apparatus

Participants sat in a dimly lit room in front of a computer monitor at a viewing distance of 1 m. 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 15.5 cd/m². One eye viewed the target stimuli, the other eye the masks.

The target stimuli were three grey discs (Ø 0.29°, mean luminance: 25.4 cd/m²), located in the centre and 5.72° visual angle to the left and right of a grey fixation dot. On each trial, the color of exactly one of the targets changed to white (mean luminance: 224.2 cd/m²) for a duration of 100 ms. This change in brightness will be referred to as 'flash'. To suppress the flash's perceptual visibility, the other eye was shown three dynamic Mondrians (Ø 2°, mean luminance: 35.6 cd/m²) (Maruya et al., 2008; Tsuchiya & Koch, 2005). We employed dynamic continuous flash suppression, as this proved a powerful and reliable method to suppress perceptual awareness of a brief and hence relatively salient flash. To match the target's location the Mondrians' were also located in the centre or 5.72° to the left and right of the fixation dot. Each Mondrian consisted of sinusoidal gratings (Ø 0.57°) which changed their color and position randomly at a frequency of 10 Hz. Each grating's texture was shifted every 16.6 ms to generate apparent motion. Visual stimuli were presented with a fixation spot in the centre of the screen and were framed by a grey, isoluminant square apperture of 8.58° x 13.69° in diameter to aid binocular fusion.

Auditory stimuli were pure tones with a carrier frequency of 1 kHz and a duration of 100 ms. They were presented via four external speakers, placed above and below the monitor. Upper and lower speakers were aligned vertically and located 2.3° to the left and 2.3° to the right of the monitor's centre. The volume of the individual speakers was matched and the height of the chinrest was set so participants' ears were level with the horizontal midline of the monitor. This made sure, that participants perceived the sounds coming from the horizontal midline of the monitor, co-located with the visual signals. Speakers' horizontal location was chosen by trading off physical alignment of visual and auditory stimulus locations and sound localization performance. Moreover, it traded off optimization for the two research questions we addressed in this study: (i) the role of audiovisual localization and (ii) auditory bias on perceived visual location. At a distance of 2.3° mean sound localization accuracy amounted to ~70%.

Psychophysical stimuli were generated and presented on a PC running Windows XP using the Psychtoolbox version 3 (Brainard, 1997; Kleiner, Brainard, & Pelli, 2007) running on Matlab 7 (Mathworks, Nantucket, Massachusetts). 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 (GeForce 8600GT graphics card). Auditory stimuli were digitized at a sampling rate of 44.8 kHz via an M-Audio Delta 1010LT sound card and presented at a maximal amplitude of 73 dB sound pressure level. Exact audiovisual onset timing was confirmed by recording visual and auditory signals concurrently with a photo-diode and a microphone.

Experimental design

Participants were presented with an auditory beep emanating from either the left or right. In synchrony with the beep, one eye was presented with a brief flash in the centre or participants' left or right hemifield. The visibility of the 'flash' was suppressed by presenting masks to the other eye using the method of dynamic continuous flash suppression (Maruya et al., 2008). Hence, the 3 x 2 factorial design manipulated (1) 'flash location' (3 levels: left, centre, right) and (2) 'sound location' (2 levels: left, right) (Figure 4.1 A). On each trial, participants located the flash (left, right or centre). Moreover, they performed a graded detection task by judging the visibility of the flash (invisible, unsure, visible).

This experimental design enabled us to address two questions: First, we investigated whether participants were better at detecting the flash, when auditory and visual signals were approximately collocated. Second, as the flash was visible only in a fraction of trials, we were able to quantify the effect of sound on localizing physically identical flashes that were visible or invisible.

Experimental procedure

As seen in Figure 4.1 B, each trial started with the presentation of the fixation dot for a duration of 1000 ms. Next, participants' one eye was presented with three grey discs, located in the centre, 5.72° visual angle to the left and right of a grey fixation dot. Participants' awareness of these discs was suppressed by showing dynamic Mondrians at the corresponding locations to the other eye (i.e. dynamic continuous flash suppression). The Mondrian masks and the discs were presented on the screen until participants had responded to all questions. The assignment of eyes to disks or masks was changed after each trial, to enhance suppression. After a random interval

of 500-1000 ms one of the three discs 'flashed', i.e. changed its luminance for a duration of 100 ms. In synchrony with the flash, an auditory beep was played from the left or right. In addition, on 22.2% of the trials, the so-called catch trials, participants were also asked to locate the sound (left vs. right discrimination; in addition to the visibility judgment and flash localization). This allowed us to assess the spatial information that is available for sound localization. Moreover, it ensures that participants did not completely ignore the sound.

Participants responded by pressing one of three buttons on a keyboard. The button assignment was counterbalanced across participants as follows: Participants used three sets of buttons to respond to the three question types (flash localization, sound localization (on catch trials only) and visibility judgment). Each set contained three buttons, one central, one to the left and one to the right. One set of buttons was operated with one hand and the other two sets were operated with the other hand. The association of the hands to the button sets was counterbalanced across participants. Moreover, we also counterbalanced the button response assignment for the flash visibility question. Within subjects we counterbalanced the two possible question orders (i.e. i. flash localization, ii. sound localization (only on catch trials), iii. visibility judgment; alternatively: i. sound localization (only on catch trials), ii. flash localization, iii. visibility judgment).

Prior to the main experiment, participants were familiarized with stimuli and task. First, they completed 2-3 sessions of sound localization. Next, there were two short practice sessions of the main paradigm. During the main experiment participants completed a total of 24 experimental sessions distributed over two successive days, resulting in a total of 1296 trials (i.e. 216 trials per condition).

Analysis

Our analysis addressed two questions:

Effect of spatial congruency on visibility judgment

We investigated whether a synchronous sound boosts 'a suppressed visual signal' into participants' awareness depending on spatial congruency. In other words, we asked whether participants were better at detecting a flash, when the sound was approximately collocated with the flashing disc. Visibility judgment as the dependent variable was quantified as the percentage of non-catch trials judged as visible. As participants' visibility judgment depended on stimulus eccentricity (central flashes are projected on the fovea centralis which is responsible for central sharp vision, hence it is expected that detection performance will be higher here than in the periphery), we limited this analysis only to those trials with left / right flashes and excluded trials with flashes in the centre. Moreover, we pooled over the left and right hemifield as there was no significant difference between left and right hemifield in percentage judged visible. Hence, congruent conditions included flash left / sound left and flash right / sound right combination. Likewise, incongruent conditions included flash left / sound right and flash right / sound left combinations. We performed paired t-tests to compare participants' visibility judgment between congruent and incongruent conditions. However, to be consistent with the statistical analyses used for comparisons concerning the relative auditory weight (detailed in the next paragraph) we also performed a non-parametric bootstrap test based on the one-sample tstatistic for the congruent minus incongruent difference (Johnson, 2001).

Effect of sound location on perceived flash location as a function of visibility

We investigated whether the influence of the sound on flash localization depended on the visibility of the flash. Critically, the flash signal intensity was fine-tuned in several pilot studies, so that approximately 50% of the flashes were judged invisible across participants at the group level. Hence, the flash visibility varied across trials and participants because of internal systems noise and participant-specific effects rather than external signal strength. We hypothesized that the influence of the true sound location would be inversely related to flash visibility. In other words, we expected that the influence of the sound on perceived flash location should be maximal for trials where the flash was judged invisible.

To quantify the influence of true sound location on participants' perceived flash location, we first coded the perceived and true flash and sound locations as -1 for left, 0 for centre and 1 for right. Separately for visible, unsure and invisible trials, we then estimated a general linear model where participants' perceived flash location as the dependent variable was predicted by the true flash and sound location on each trial:

$$V_p = \beta_0 + (\beta_V * V_t) + (\beta_A * A_t) + \epsilon$$

with V_p = perceived/reported flash location, V_t = true flash location, A_t = true sound location, β_0 = intercept term, β_V = coefficient for true flash location, β_A = coefficient for true sound location, ε = error term. As the audiovisual spatial discrepancies in this experiment were smaller than 10° visual angle, we assumed that auditory and visual signals are combined linearly as assumed under the standard forced fusion model (Alais & Burr, 2004). In other words, the influence of the true sound location (as
quantified by the regression coefficient β_A) is assumed not to vary with the spatial discrepancy. Hence, we did not include an interaction term $A_t x V_t$ in the regression model.

We computed the relative auditory weight as an index of the influence of sound on perceived flash location according to:

Relative Auditory Weight
$$= rac{oldsymbol{eta}_A}{oldsymbol{eta}_A + oldsymbol{eta}_V}$$

We tested whether the relative auditory weight was greater than zero using onesample t-tests. A positive auditory weight indicates that the perceived visual location is shifted towards the true auditory location as expected for a reverse ventriloquist illusion. A negative auditory weight suggests that the perceived visual location is shifted away from the true auditory location (i.e. repulsion effect). An auditory weight that is not significantly different from zero suggests that the location of the sound does not significantly influence the perceived location of the flash. For comparison across visibility levels a one-way repeated measures ANOVA was performed with factor visibility. Planned pairwise comparisons were performed using paired t-tests. Moreover, to refrain making any parametric assumptions (n.b. the relative auditory weight conforms to a ratio distribution) we repeated these comparisons using nonparametric bootstrap-based tests.



Figure 4.2 Behavioural results. (A) Bar plots showing the percentage of flashes judged visible, unsure and invisible for audiovisual spatially congruent and incongruent conditions (across subjects' mean \pm SEM). Critically, the % judged visible was significantly higher for audiovisual spatially congruent relative to incongruent conditions. (B) Violin plot showing the distribution of the individual differences in percentage of flashes judged visible between the spatially congruent and incongruent conditions. The individual data points are overlaid. (C) Bar plots showing the relative auditory weights (across subjects' mean \pm SEM) obtained from the regression model separately for visible, unsure and invisible trials. As the regression model (specified in the methods) can only be estimated with at least three trials present for a particular visibility level, the number of subjects varies across the different visibility levels (visible: n = 21; unsure: n = 22; invisible: n = 23).

Results

Effect of spatial congruency on visibility judgment

Figure 4.2 A shows the percentage of trials judged visible, unsure and invisible. As expected we observed a significant increase in percentage judged visible, when the sound was presented in the same relative to the opposite hemifield (percentage judged visible: congruent – incongruent (mean \pm SEM): 1.8 \pm 0.51; Cohen's d: 0.73; paired-samples t-test, t(22) = 3.51, p = 0.002, bootstrap-based p < 0.001; Figure 4.2

B). Conversely, we observed a significant decrease in percentage judged invisible for spatially congruent relative to incongruent trials (percentage judged invisible: congruent - incongruent (mean ± SEM): -1.94 ± 0.65; Cohen's d: -0.62; pairedsamples t-test, t(22) = -2.98, p < 0.007; bootstrap-based p = 0.011). This suggests that a sound influences whether visual signals reach perceptual awareness depending on audiovisual spatial congruency. As we did not include any trials where no flash was presented, we cannot compute the d-prime for the congruent and incongruent conditions or formally dissociate sensitivity and decisional bias. However, as the evaluation of audiovisual spatial congruency obviously entails spatial localization of both flash and sound, it is inconsistent to assume that audiovisual spatial congruency takes effect by influencing the decisional bias in the visibility judgment task. Moreover, had we included trials without a flash to estimate the false alarm rate, we would have still included the same false alarm rate for spatially congruent and incongruent conditions when computing the d-prime. In other words, the % judged visible directly corresponds to the d-primes for congruent and incongruent conditions.

Effect of sound location on perceived flash location as a function of visibility

We quantified the influence of sound on perceived flash location across visibility levels in terms of the relative auditory weight obtained from the regression approach (see methods). As the regression model specified can only be estimated when at least three trials are present for a particular visibility level, the relative auditory weights are based on a different number of subjects across the different visibility levels (visible: n = 21; unsure: n = 22; invisible: n = 23). Figure 4.2 C shows the

relative auditory weights on the perceived location of a visible, unsure and invisible flash. We observed positive relative auditory weights for all three visibility levels. Critically, the relative auditory weights significantly differed across visibility levels (main effect of visibility: F(1.6,29.8) = 25.6, MSE = 3.75, p < 0.001). More specifically, the relative auditory weight for visible trials was significantly different from that for unsure or invisible trials (paired-t test: unsure-visible t(19) = 6.54, parametric p < 0.001, bootstrap-based p < 0.001; invisible-visible t(20) = 6.44, parametric p < 0.001, bootstrap-based p < 0.001; n.b. the degrees of freedom vary as different numbers of subjects could be included, see above). As expected the auditory influence on perceived flash location was greatest when the flash was judged invisible.

Discussion

Combining spatial audiovisual stimulation and continuous flash suppression we investigated whether and how signals from different sensory modalities can interact prior to perceptual awareness. Continuous flash suppression is thought to affect visual perception by attenuating neural activity already in primary visual cortices similar to reducing the contrast of the stimulus (Yuval-Greenberg & Heeger, 2013). Essentially it increases the internal noise in the primary visual cortex, reducing the signal-to-noise ratio of the signal. It is likely that this attenuation of neural activity destabilizes neural representations and prevents them from propagating up the cortical hierarchy thereby obliterating them from perceptual awareness. To measure the effect of a concurrent sound on participants' visual awareness, we tuned the strength of the visual flash such that it entered participants' awareness only on a fraction of trials. We then investigated whether the effect of a synchronous sound on

participants' visibility judgment depended on audiovisual spatial congruency. Indeed, our results demonstrate that participants were more likely to detect the flash, when the sound was co-localized than non-collocated with the flash. In support of an 'attention-free' account of audiovisual integration these results suggest that an aware auditory signal can boost a weak visual signal into participants' awareness. Critically, the sound was brief and synchronous with the flash across all conditions. Hence, the effects of spatial congruency are unlikely to be explained by a reduction in temporal uncertainty or more precise temporal expectations. Instead they suggest that audiovisual interactions prior to perceptual awareness are governed not only by temporal (as shown by Alsius and Munhall, 2013) but also by spatial constraints. There are at least two mechanisms by which a collocated sound may enhance flash visibility. First, a collocated sound may influence visual perception via bottom-up mechanisms that boost visual salience and enable the formation of spatially more precise salience maps. Second, a collocated sound may reduce visual spatial uncertainty via top-down mechanisms that enable more effective allocation of attentional resources and stabilize visual representations potentially even after they have accessed awareness. In the current paradigm, top-down mechanisms may be less likely because audiovisual signals were presented in synchrony, therefore a colocated sound could not direct observers' attention prior to stimulus presentation. Yet, future electrophysiological studies are needed to determine the role of bottom-up from top-down mechanisms in audiovisual interactions during flash suppression.

In sum, our results suggest that audiovisual interactions emerge largely prior to awareness governed by the classical principles of spatial congruency (Stein & Meredith, 1993; Wallace et al., 2004). These interactions in turn enhance stimulus

salience and thereby enable a visual signal to elude flash suppression and enter participants' awareness. A controversial guestion is whether spatial congruency acts as a fundamental principle of multisensory integration or depends on stimulus characteristics and task-constraints (for excellent review see Spence, 2013). Accumulating evidence from behavioural research suggests that spatial congruency benefits performance predominantly in tasks where spatial information is relevant (e.g. overt or covert spatial orienting - Arndt & Colonius, 2003; Diederich, Colonius, Bockhorst, & Tabeling, 2003; Harrington & Peck, 1998; Santangelo & Spence, 2008; Spence, 2010), but less so in detection (e.g. redundant target paradigms or identification tasks – Bertini, Leo, & Làdavas, 2008; Forster, Cavina-Pratesi, Aglioti, & Berlucchi, 2002; Girard, Collignon, & Lepore, 2011). The current study cannot fully exclude that the role of spatial congruency emerges because subjects were engaged in both visibility judgment and spatial localization. Yet, as in previous masking studies (e.g. Bolognini et al., 2005; Frassinetti et al., 2002) an increase in detection performance was also observed in the absence of an additional localization task, spatial task demands do not seem absolutely critical. Instead, we would suggest that concurrent sounds automatically interact with visual signals as a function of spatial discrepancy in low level visual areas thereby amplifying the neural activity and boosting the flash into participants' awareness. Future studies are needed to further characterize the critical spatial integration window by systematically manipulating the spatial discrepancy of the audiovisual signals under flash suppression. Together with additional EEG and fMRI studies this research line would allow us to further pinpoint the cortical level at which sounds interact with visual processing under flash suppression.

In addition to judging the flash's visibility participants also located the flash on each trial. As the spatial discrepancy was approximately 8 degrees visual angle, we would expect that a concurrent, yet spatially discrepant sound biases the perceived visual location (Alais & Burr, 2004). The critical question of this study was whether participants' perceived flash location was influenced by the sound as a function of flash visibility. As expected we observed that the influence of sound location on perceived flash location increased gradually from visible to unsure and invisible trials. This audiovisual spatial bias profile is consistent with the principle of reliabilityweighted integration where a stronger weight should be given to the more reliable signal. Indeed, numerous psychophysics and recent neurophysiological studies (Alais & Burr, 2004; Ernst & Banks, 2002; Christopher R. Fetsch, Deangelis, & Angelaki, 2013; Christopher R. Fetsch, Pouget, Deangelis, & Angelaki, 2012; Morgan, DeAngelis, & Angelaki, 2008) have demonstrated that humans and nonhuman primates integrate signals weighted by their reliability approximately in accordance with predictions from Maximum Likelihood Estimation. In contrast to these previous studies we did not manipulate the reliability of the external signals. Instead, the flashes were physically identical across all visibility levels. Yet, identical physical signals will elicit neural representations that vary in their reliability across trials because of trial-specific internal systems noise (Faisal, Selen, & Wolpert, 2008). Thus, as the brain does not have access to the true physical reliability of the sensory signals but only to the uncertainty of the internal representations, it is likely that the sensory weights in the integration process depend on both the noise in the environment and the trial-specific noise in the neural system. Thus, our findings suggest that the relative auditory weight in the integration process depends on the

reliability of the trial-specific internal representation evoked by the visual signal. For example, if the visual signal is too weak to elude flash suppression and propagate to higher order association areas, 'multisensory' representations for instance in parietal areas or response selection processes in frontal areas may be more strongly dominated by auditory inputs (Bisley & Goldberg, 2010; Gottlieb, Kusunoki, & Goldberg, 1998; Emiliano Macaluso & Driver, 2003, 2005). As sensory noise also determines flash visibility, one may also argue that visible flashes bias participants' perceived sound location via higher order cognitive biasing mechanisms. In other words, if a flash elicits a noisy representation that does not enter participants' awareness, participants locate the sound purely based on the auditory input. By contrast, if a flash elicits a strong sensory representation that enters awareness, participants' perceptual decision is biased by the concurrent visual input. As we did not include any trials where no flash was presented, we cannot compute the d-prime for the visible, unsure and invisible conditions or formally dissociate sensitivity and decisional bias. Future neurophysiological and neuroimaging studies are required to determine the neural mechanisms underlying this reliability weighting that emerges from internal noise rather than manipulation of external signal strength.

CHAPTER 5: TO INTEGRATE OR NOT TO INTEGRATE: TEMPORAL DYNAMICS OF BAYESIAN CAUSAL INFERENCE

This study is a result of collaboration between Uta Noppeney and Máté Aller and is being prepared for publication as a research article. The study was designed by UN and MA, data were collected by MA, analysed by MA (supervised by UN). The manuscript was written by MA building upon the methods of previous studies by Rohe and Noppeney (Rohe & Noppeney, 2015a, 2016).

Introduction

In our natural environment we are exposed to a constant inflow of sensory signals: We hear sounds, see objects, touch surfaces and smell the air we breathe in. How the brain constructs a veridical representation of the environment from this sensory cacophony, seemingly effortlessly, remains unclear. The brain needs to solve two fundamental computational problems: First, it must determine whether sensory inputs originate from common or independent sources, the so-called causal inference problem. Spatio-temporal correspondences between the sensory stimuli as well as prior knowledge about the environment inform the brain whether signals are likely to come from a common source. Second, signals should be integrated or segregated depending on the most likely causal structure: if they are caused by a common source the brain can use the redundant pieces of information to form the most reliable integrated percept. Critically however, signals from independent sources should be segregated.

Hierarchical Bayesian Causal Inference provides a unifying framework to incorporate both sensory integration and segregation in perception: it does so by explicitly

modelling the two possible causal structures (i.e. common or independent) that could give rise to the sensory inputs. In line with Helmholtz' notion of 'unconscious inference' (von Helmholtz, 1896), the brain is then thought to invert this generative model during perception. In case of a common source, signals are integrated weighted by their relative sensory reliabilities. In case of independent sources, they are segregated and processed separately. Critically, the brain has only access to noisy sensory information about the environment, including its causal structure. To account for the uncertainty, the estimates (e.g. the object's location) under the two causal structures are weighted in the final estimate by their causal structure's posterior probability (for other decisional strategies see Wozny et al. (2010)).

A solid body of behavioural evidence suggests that human observers combine sensory signals near-optimally in line with the predictions of Bayesian Causal Inference (Körding et al., 2007; Rohe & Noppeney, 2015b; Wozny et al., 2010). Most prominently, during spatial orientation and localization observers seamlessly transition from sensory integration to segregation depending on spatial disparity between the individual stimuli. When audiovisual signals are in close proximity they are integrated, which leads to strong crossmodal spatial biases operating from vision to audition and vice versa depending on their relative sensory reliabilities; however, when they are far apart, signals are unlikely to come from a common source, hence integration breaks down and the crossmodal biases are weakened. A recent study combined fMRI, multivariate pattern decoding and Bayesian Causal Inference in a spatial localization task and demonstrated that representations of multisensory perception and Bayesian Causal Inference develop across a hierarchy of brain areas (Rohe & Noppeney, 2015a). At the bottom of the hierarchy, auditory and visual

sensory areas represent the signals assuming that they originate from independent sources (i.e. full segregation). At an intermediate stage, the representations in posterior parietal sulcus reflect signals as they were caused by a common source (i.e. forced fusion). The uncertainty about the world's causal structure is taken into consideration only at the top of the hierarchy in anterior parietal cortex, where estimates of the lower hierarchical levels are combined weighted by their causal probability (i.e. Bayesian Causal Inference). Thus, Bayesian Causal Inference recruits entire cortical hierarchies that encode different spatial estimates. These results were further corroborated by a different analysis approach (Rohe & Noppeney, 2016), which does not assume the predictions of the Bayesian Causal Inference model, rather computes the relative weight each sensory modality is given in the neural representations. Comparing the relative weights between the various experimental conditions across the cortical hierarchy gives a 'model free' view on how sensory stimuli are integrated in the brain. Primary sensory areas showed limited integration depending on spatial disparity, whereas parietal areas reflected sensory estimates weighted by their sensory reliabilities and task relevance. Furthermore, integration degraded in parietal cortices when signals were far apart (i.e. less likely to originate from a common source). Collectively these results further strengthen the hypothesis that multisensory integration evolves across a hierarchy of cortical areas, each reflecting distinct computational principles. Yet, given the limited temporal resolution of fMRI, this previous study was blind to the within-trial temporal dynamics of multisensory perception and Bayesian Causal Inference. How does it dynamically arbitrate between sensory integration and segregation when presented

with audiovisual signals that vary in their audiovisual spatial disparity? How do spatial estimates evolve dynamically?

To determine how the brain forms multisensory representations over time at millisecond resolution the current study combined EEG, multivariate pattern decoding and models of Bayesian Causal Inference. Human observers were presented with audiovisual signals that varied in their spatial disparity and relative sensory reliability and reported either the auditory or the visual stimulus location. Linear support vector machine models were trained on the spatially congruent conditions and generalized to spatially congruent and incongruent conditions. Then the decoded stimulus locations were used to determine the temporal dynamics of multisensory perception using a 'model-free' (relative audiovisual weight) and a 'model-based' (i.e. Bayesian Causal Inference) method.

Methods

Participants

Sixteen right-handed participants gave informed consent to participate in the experiment and received monetary compensation. Three participants dropped out from the experiment during the first session: two participants were excluded because they blinked excessively (because of dry eyes) and one participant withdrew from the experiment. The remaining thirteen participants (7 females, mean age = 22.1; SD = 3.0) completed the experiment. All participants had no history of neurological or psychiatric illnesses, had normal or corrected-to-normal vision and normal hearing. The study was approved by the research ethics committee of the University of

Birmingham and was conducted in accordance with the principles outlined in the Declaration of Helsinki.

Stimuli and apparatus

The visual (V) stimulus was a cloud of 20 white dots (diameter = 0.43° visual angle) sampled from a bivariate Gaussian distribution with vertical standard deviation of 2° and horizontal standard deviation of 2° or 12° visual angle presented on a dark grey background (67% contrast) at a frame rate of 60 Hz for 50 ms in synchrony with the auditory stimulus. The auditory (A) stimulus, was a 50 ms long burst of white-noise with 5 ms on/off ramp. Stimuli were presented using Psychtoolbox version 3.0.11 (Brainard, 1997; Kleiner et al., 2007) under MATLAB R2014a (MathWorks Inc.) on a desktop PC running Windows 7. Participants were seated at a desk resting their head on a chin rest located at a distance of 475 mm from a gamma-corrected 30" LCD monitor with a resolution of 2560 x 1600 pixels. Auditory stimuli were delivered through eight external speakers mounted on the top and bottom edges of the monitor. Upper and lower speakers were aligned vertically in pairs and placed at the appropriate horizontal stimulus locations. The volume of the individual speakers was matched within \pm 0.2 dB SPL verified by 3 individual measurements with a sound level meter (CEL24x, NoiseMeters Ltd., Ockley, UK). The height of the chinrest was set so participants' ears were level with the horizontal midline of the monitor. This made sure, that participants perceived the sounds coming from the horizontal midline of the monitor. Participants responded by pressing one of four response buttons on a USB keypad with their index, middle, ring and little finger respectively.

Experimental design and procedure

In a spatial ventriloquist paradigm, participants were presented with synchronous, yet spatially congruent or discrepant visual and auditory signals (Figure 5.1A). On each trial, visual and auditory locations were independently sampled from four possible locations along the azimuth (i.e., -10°, -3.3°, 3.3° or 10°) leading to four levels of spatial disparity (i.e., 0°, 6.6°, 13.3° or 20°). In addition, we manipulated the reliability of the visual signal by setting the horizontal standard deviation of the Gaussian cloud to 2° (high reliability) or 12° (low reliability) visual angle. In an inter-sensory selectiveattention paradigm, participants either reported their auditory or visual perceived stimulus location. Hence, the 4 x 4 x 2 x 2 factorial design manipulated (1) the location of the visual stimulus ([-10°, -3.3°, 3.3°, 10°], i.e., the mean of the Gaussian) (2) the location of the auditory stimulus ([-10°, -3.3°, 3.3°, 10°]) (3) the reliability of the visual signal ([2°,12°], standard deviation of the Gaussian) and (4) task-relevance (auditory- / visual-selective report). Trials were organized in runs of 128 with short breaks between them. Within each run, the responded modality (auditory vs. visual) and the responding hand (left vs. right) was kept constant, counterbalanced within and between participants. The rest of the conditions were randomized and equally likely to appear within each run. Each experimental condition was presented 96 times, resulting 6144 trials in total.

On each trial, synchronous auditory and visual stimuli were presented for 50 ms, which was followed by a response cue 1000 ms after stimulus onset (Figure 5.1B). The response cue consisted of a central pure tone (1000 Hz) and a blue colour change of the fixation cross presented synchronously for 100 ms. Participants were instructed to withhold their responses and to avoid blinking until the cue appeared.

The next stimulus was presented after response/fixation period of variable length (2.6-3.1 s).

Participants performed the experiment over the course of 3 days. The first experimental day started with practice runs to familiarize participants with both visual and auditory response runs. On the remaining days participants performed only one short practice run with auditory localization as it was the more challenging task. Participants were instructed to fixate on a central fixation cross throughout the experiment.

Eye movement recording and analysis

To address potential concerns that results may be confounded by eye movements, we recorded the eye movements of the participants. Eye recordings were calibrated in the recommended field of view (32° horizontally and 24° vertically) for the EyeLink 1000 Plus system with the desktop mount at a sampling rate of 2000 Hz. Eye position data were on-line parsed into events (saccade, fixation, eye blink) using the EyeLink 1000 Plus software. The 'cognitive configuration' was used for of saccade detection (velocity threshold = 30° /sec, acceleration threshold = 8000° /sec², motion threshold = 0.15°) with an additional criterion of radial amplitude larger than 1°. Individual trials were rejected based on events occurring from -100 to 700 ms post-stimulus.



Figure 5.1 Experimental design, example trial, and behavioural audiovisual weights (A) The experiment manipulated (1) the location of the visual (V) signal (-10°, -3.3°, 3.3°, and 10°), (2) the location of the auditory (A) signal (-10°, -3.3°, 3.3°, and 10°), (3) the reliability of the visual signal (high [VR+] versus low [VR-], as defined by the spread of the visual cloud), and (4) task relevance (auditory versus visual report). (B) Time course of an example trial. (C) Schematics of the models used to explain behavioural and neural responses: 'segregation auditory' model (SegA), 'segregation visual' model (SegV), 'forced fusion' model (Fusion) and the 'Bayesian Causal Inference' model (with model averaging) (BCI).

EEG data acquisition

Continuous EEG signals were recorded from 64 channels using Ag/AgCI active electrodes arranged in 10-20 layout (ActiCap, Brain Products GmbH, Gilching, Germany) at a sampling rate of 1000 Hz. Channel impedances were kept below $10k\Omega$.

General analysis strategy

To characterize how human observers integrate auditory and visual signals into spatial representations at the behavioural and neural level, we developed a common

analysis strategy for both (i) reported auditory and visual spatial estimates (i.e., participants' behavioural localization responses) and (ii) the spatial estimates decoded from the time course of EEG responses. We used two main analysis approaches similarly to Rohe and Noppeney (Rohe & Noppeney, 2015a, 2016): The first approach we call 'model free' as it does not depend on predictions of a specific model of sensory integration. The second approach we call 'model based' as it is based on the predictions of the Bayesian Causal Inference model of sensory integration (Körding et al., 2007).

In the 'model free' analysis approach we quantified the influence of the auditory and visual signals on the reported (behavioural) or decoded (neural) auditory or visual spatial estimates using a linear regression model. In this regression model, the perceived/decoded spatial locations were predicted by the true spatial locations separately for auditory and visual signals for each of the eight conditions in the 2 (visual reliability: high vs. low) x 2 (task-relevance: auditory vs. visual report) x 2 (spatial disparity: $\leq 6.6^{\circ}$ vs. > 6.6°) factorial design (Figure 5.2). Hence, the regression model included 16 regressors in total, i.e., 8 (conditions) x 2 (true auditory or visual spatial locations). The auditory (β_A) and visual (β_V) parameter estimates quantified the influence of auditory and visual signals on the perceived/decoded signal location for a particular condition. To obtain a summary index for the relative audiovisual weights, we computed the relative audiovisual weight (w_{AV}) as the fourquadrant inverse tangent of the visual (B_V) and auditory (B_A) parameter estimates from the regression model for each of the eight conditions based on participants' behavioural responses (Figure 5.3) and their decoded spatial estimates across the time course of the trial (Figure 5.6A-D). If the reported/decoded estimate is

dominated purely by the visual signal, w_{AV} is 90°. For pure auditory dominance, it is 0°.

In the 'model based' analysis approach we modelled the reported (behavioural) or decoded (neural) spatial estimates according to the Bayesian Causal Inference model and its constituent models. Specifically, we fitted four models (Figure 5.1C): the 'segregation auditory' and 'segregation visual' models assume only one auditory (resp. visual) cause and take only the auditory (resp. visual) signal into consideration to model the response. The 'forced fusion' model assumes a single cause and weights the signals according to their relative reliabilities for the final estimate, equivalent to the Maximum Likelihood Estimation (Ernst & Banks, 2002). The 'Bayesian Causal Inference' (BCI) model estimates the posterior probability of the causal structure underlying the stimuli (common vs separate causes) and weights the fusion and segregation estimates according to the posterior probability of their causal structure (i.e. with model averaging, see Körding et al. (2007) for details of the model fitting process). Hence the BCI model incorporates the other tested models.

Behavioural analysis

In the 'model free' approach we calculated the behavioural w_{AV} from participants' responses to assess the influence of the auditory and visual stimuli on the perceived stimulus location (Figure 5.3A). The main effects of visual reliability, task-relevance and spatial disparity on the behavioural w_{AV} were tested using the parametric Watson-Williams test. Two-way interactions were tested using the Harrison-Kanji test (we used the CircStat toolbox for the statistical analyses (De Leeuw, 2009)).

In the 'model based' approach the four models ('segregation auditory', 'segregation visual', 'forced fusion' and 'BCI', see General analysis strategy section) were fitted to the behavioural responses. The model fit was evaluated in terms of the Bayesian Information Criterion (BIC = LL - 0.5 M ln(N), LL = log-likelihood, M = number of parameters, N = number of data points (T. Watanabe & Flavin, 1976)) separately for each subject. BIC penalizes model performance based on model complexity. Then we performed Bayesian model selection (Rigoux et al., 2014) at the group level as implemented in SPM12 (Friston, Holmes, et al., 1994) to compute the exceedance probability of each model (i.e. the probability that the model at hand is more likely than any of the others given the data).

EEG data analysis

Pre-processing

Pre-processing was performed with the FieldTrip toolbox (Oostenveld, Fries, Maris, & Schoffelen, 2011). Raw data were high pass filtered at 0.1 Hz, re-referenced to average reference, and low pass filtered at 120 Hz. Trials were extracted with 100 ms pre-stimulus and 700 ms post-stimulus period, were baseline corrected, temporally smoothed with a 20 ms moving window and down-sampled to 200 Hz (note, that a 20 ms moving average is comparable to a Finite Impulse Response (FIR) filter with a cut-off frequency of 50 Hz). Trials containing artefacts were rejected based on visual inspection. Furthermore, trials were rejected if they (i) contained eye blinks, or (ii) saccades, or if (iii) the eye gaze was away from the fixation cross further than 2 degrees, or if (iv) a response was made prior to the response cue or (v) there was no response.

Multivariate decoding

We trained linear support vector regression models (SVR, LIBSVM (Chang & Lin, 2011)) separately for each time point up to 700 ms post stimulus. SVR models were trained in a 10-fold stratified cross-validation procedure. First, data from all conditions were randomly split into 10 folds making sure all spatial locations from all conditions are equally well represented in each fold. The SVR model was trained to learn the mapping from activation patterns to space using examples from the audiovisual congruent conditions from all but one fold. Then the learnt mapping was used to decode the spatial location from the evoked potential patterns of the audiovisual congruent and incongruent examples of the remaining fold. This procedure was repeated for all folds. The hyper parameters of the SVRs were fixed to their default values as defined in LIBSVM (C = 1, ν = 0.5). To maximize the signal to noise ratio evoked potentials were first computed by averaging over sets of 8 randomly assigned individual trials from the same condition (Grootswagers et al., 2017).

First, we examined if the true audiovisual stimulus locations could be successfully decoded from the EEG response patterns over the course of the trial. Decoding accuracy was evaluated as the Pearson correlation coefficient between the true and decoded auditory locations on audiovisual congruent trials (Figure 5.5A). To assess whether the decoding accuracy was better than chance anywhere over the time course we entered the Fisher transformed Pearson correlation coefficients into a between subjects permutation test based on the one sample t-statistic (as implemented in the FieldTrip toolbox (Maris & Oostenveld, 2007; Oostenveld et al., 2011)). To correct for multiple comparisons a cluster correction method was used

based on the maximum of the sums of cluster level statistics ('maxsum') (Maris & Oostenveld, 2007; Nichols & Holmes, 2003)

Second, we investigated the temporal dynamics of how auditory and visual signals are integrated into spatial representations. For this, we focused on the spatial locations decoded by the SVR procedure from the spatially congruent and incongruent conditions that provide information about how the brain at a given time point combines visual and auditory spatial signals into spatial representations. Using the same regression analysis approach as for behavioural localization responses, we quantified the influence of the true auditory and visual signal location on the decoded spatial estimates separately for each time point (see General analysis strategy section). We computed the 'neural' relative audiovisual weight w_{AV} as the angle between the auditory and visual parameter estimates of the linear regression for each time point and bootstrapped confidence intervals of mean w_{AV} (Figure 5.6A-D).

Permutation of circular indices w_{AV} for behavioural and neural data

We performed the statistics on the behavioural and neural audiovisual weight indices using a two (auditory vs. visual report) x two (high vs. low visual reliability) x two (large vs. small spatial disparity) factorial design based on the likelihood ratio statistics for circular measures (LRTS) (Anderson & Wu, 1995). Similar to an analysis of variance for linear data, LRTS computes the difference in log-likelihood functions for the full model that allows differences in the mean locations of circular measures between conditions and the reduced null model that does not model any mean differences between conditions. LRTS were computed separately for the main effects and interactions. To refrain from making any parametric assumptions, we evaluated the main effects of visual reliability, task-relevance, spatial disparity and their

interactions in the factorial design using permutation testing of the LRTS. Permutations were constrained to occur within each participant. For the main effects of visual reliability, task-relevance and spatial disparity, w_{AV} values were permuted within the levels of the non-tested factors. For tests of the two-way interactions, we permuted the simple main effects of the two factors of interest within the levels of the third factor (Edgington & Onghena, 2007). For tests of the three-way interaction, values were freely permuted across all conditions (Gonzalez & Manly, 1998). All tests were based on 5000 random permutations and a right-tailed p-value was computed.

To assess the similarity between behavioural and neural audiovisual weights (w_{AV}), we computed the circular correlation coefficient (as implemented in the CircStat toolbox (De Leeuw, 2009)) between the 8 neural and 8 behavioural w_{AV} s from our 2 (high vs. low visual reliability) x 2 (auditory vs. visual report) x 2 (large vs small spatial disparity) factorial design separately for each time point.

To correct for multiple comparisons within the one (i.e. time) data, cluster-level inference was used based on the maximum of the summed t-values within each cluster ('maxsum') with a cluster defining threshold of p < 0.05 and a two-tailed p-value was computed.



D+/D- X VR+/VR- X A/V report

Figure 5.2. The audiovisual weight index. Audiovisual weight index wAV was computed as the angle between the auditory and visual regression coefficients. Audiovisual weight index wAV as a function of audiovisual spatial disparity (small [\leq 6.6°, D-] versus large [>6.6°, D+]), task relevance (auditory versus visual report), and visual reliability (high [VR+] versus low [VR-]) are shown

Results

Behavioural results

First, we evaluated the behavioural relative audiovisual weight w_{AV} ('model free' analysis) (Figure 5.3A). We found, that the visual signal location influenced participants' percept depending on the reliability of the visual signal: it was greater for high reliability versus low reliability (main effect of visual reliability: p < 0.001). Similarly the relative weight was greater if the visual stimulus location was reported versus the auditory (main effect of task-relevance p < 0.001), in other words participants reported different locations for identical stimuli depending on the relevant sensory modality. Critically this difference increased as the spatial disparity increased (interaction between spatial disparity and task relevance: p < 0.001) and the reliability of the visual stimulus decreased (interaction between visual reliability and task relevance p < 0.001). In other words audiovisual integration broke down as spatial disparity increased and the reliability of the visual signal decreased.

Second, we plotted the group level response distributions for the 4 (A location) x 4 (V location) x 2 (VR+ / VR-) x 2 (A report / V report) (Figure 5.4). If participants' responses were perfect the response distributions would be impulse functions at the correct response locations, hence the width of the distributions expresses the

participants' precision. As seen in Figure 5.4 participants were more precise in locating the visual signal as compared to the auditory (blue vs red solid lines). Similarly visual localization was more precise when the visual signal was more reliable (i.e. the SD of the cloud was smaller). Importantly, visual signals biased the localization of auditory signals and this effect was critically more pronounced when the visual stimulus was more reliable or if the spatial disparity between the stimuli was small.



Figure 5.3. Behavioural audiovisual weight index. Behavioural audiovisual weight index wAV estimated from behavioural responses (A) and from the predictions of the Bayesian Causal Inference model (B; across-participants circular mean \pm 68% Cl; n = 13). The audiovisual weight index wAV is shown as a function of (i) visual reliability: high [VR+] versus low [VR-], (ii) task relevance: auditory [A] versus visual [V] report, (iii) audiovisual spatial disparity: small [\leq 6.6; D-] versus large [>6.6; D+].

Third, we formally compared how well each of the four models predicted participants' responses. The 'segregation auditory' and 'segregation visual' model assumes only one auditory (resp. visual) cause and take only the auditory (resp. visual) signal into

consideration to model the response. The 'forced fusion' model assumes a single cause and weights the signals according to their relative reliabilities for the final estimate. The 'Bayesian Causal Inference' (BCI) model estimates the posterior probability of the causal structure underlying the stimuli (common vs separate causes) and weights the fusion and segregation estimates according to the posterior probability of their causal structure, hence it incorporates the previous three models. We computed the Bayesian Information Criterion (BIC) for each model and performed Bayesian model selection, which corroborated previous findings (Körding et al., 2007; Rohe & Noppeney, 2015a), that the BCI model outperformed the other models (Exceedance probability for BCI model: 0.9999, for all other models < 0.0001). Figure 5.4 shows the group average model distributions predicted by the BCI model (dashed lines) overlaid with the behavioural responses (solid lines). Furthermore, we generated 'model responses' from the predictions of the BCI model and we subjected them to the 'model free' analysis. This enabled us to compare the profiles of audiovisual weight index w_{AV} between behavioural and 'model' responses. As shown in Figure 5.3B, the profile of the w_{AV} predicted by BCI is consistent with the profile based on behavioural responses.

Collectively both our 'model free' and 'model based' analyses confirmed, that participants integrated audiovisual signals weighted by their relative reliabilities and task relevance.



Figure 5.4 Behavioural responses and BCI model fits. Group average response distributions (solid lines) for the 4 (A location) x 4 (V location) x 2 (VR+ vs. VR-) x 2 (A response vs V response) conditions and the corresponding Bayesian Causal Inference model (with 'model averaging') predictions (dashed lines).

EEG results

Temporal dynamics of decoding accuracy

To investigate the temporal dynamics of neural representations underlying audiovisual integration we applied a support vector regression (SVR) based decoding approach to the time resolved EEG signals. First, we trained and generalized the SVR models on spatially congruent stimuli and assessed the accuracy to which the stimulus location could be decoded for each time point separately (Figure 5.5A). The decoding accuracy was expressed as the Pearson correlation coefficient between the real and decoded stimulus locations. The decoding accuracy fluctuated around chance level (R = 0) before stimulus onset and started to rise rapidly at 45 ms post stimulus. The accuracy reached a peak at 105 ms (R = 0.6) and started to decay after 210 ms post stimulus. Cluster permutation test confirmed, that the location of

the signal was successfully decoded, revealing a significant cluster starting at 55 ms post stimulus until the end of the investigated time window (p < 0.001) (Figure 5.5A shaded area).

А

0.7 Decoding accuracy AV Evoked Potential -10° 0.6 AV Evoked Potential -3.3° AV Evoked Potential 3.3° 0.5 AV Evoked Potential 10° 0.4 R 0.3 0.2 3 0.1 2 1 0 0 300 700 -1 mV 400 500 600 -100 Time (ms) -2 -3 -5 В 4 -10 -3 AV Icoation 3 10 mV 0 -2 2 4



the grand average evoked potential for the 4 congruent audiovisual [AV] conditions (pooled over visual reliabilities [VR+/-]). Shaded grey area indicates decoding performance significantly better than chance. The evoked responses are averaged across the occipital electrodes shown in the inset. (B) Grand average topographies across the 4 congruent AV conditions (pooled over VR+/-) extracted at selected timepoints illustrates the patterns upon which the decoding is based.

Temporal dynamics of neural audiovisual weight

After confirming that the spatial location of the stimuli could be successfully decoded from the EEG response patterns we investigated how visual reliability, task-relevance and spatial disparity influenced the decoded spatial locations. For this we generalized the previously trained SVR models to all trials including spatially incongruent ones as well. Applying the same approach as in the 'model free' behavioural analysis to the time resolved decoded locations we computed the neural relative audiovisual weights (w_{AV}) in our 2 (high vs. low visual reliability) x 2 (auditory vs. visual report) x 2 (large vs small spatial disparity) factorial design (Figure 5.2) for each time point separately.

First, we examined the dynamics of the effect of sensory reliability on the neural weights. As expected based on previous research (Alais & Burr, 2004; Ernst & Banks, 2002; Christopher R. Fetsch et al., 2012; Rohe & Noppeney, 2016), sensory reliability significantly influenced the neural weight as indicated by a significant cluster between 90 - 450 ms post stimulus, (p = 0.011): The neural representations reflected the visual signal location when the visual signal was reliable, however the representations shifted towards the auditory signal when visual reliability was low (Figure 5.6A). Importantly this effect did not emerge until 90 ms post stimulus which

corroborates the previous finding (Rohe & Noppeney, 2016), that sensory reliability does not influence spatial representations in early sensory areas.

Then we asked whether neural representations reflected the weighting of spatial signals according to their task relevance. The cluster based permutation test revealed a significant main effect for task relevance: one significant cluster emerged between 220 - 700 ms post stimulus (p < 0.001) (Figure 5.6B). In the visual localization task the visual signal location was given more weight relative to the auditory localization task. In other words the location of the auditory signal 'attracted' the decoded estimate when it was the task relevant modality starting from 220 ms post stimulus; once again this corroborates previous findings, where higher order areas, such as IPS0-4 showed task relevance effects (Rohe & Noppeney, 2016).

The time course of the effect for spatial disparity showed a marked difference for a short period after stimulus onset. Statistical analysis revealed a significant cluster between 45 - 150 ms post stimulus (p < 0.001). Within this time window, the visual stimulus location dominated the representations when the stimuli were in close proximity; however when they were far apart the auditory stimulus 'repelled' the decoded estimate indicated by w_{AV} values well above 90°. This indicates that the spatial window in which the stimuli occur is relevant in early multisensory processing (Figure 5.6C).

Most importantly, we also observed a significant interaction between task relevance and spatial disparity from 350 – 435 ms. As discussed in the context of the behavioural results, this interaction is a qualitative profile characteristic for Bayesian Causal Inference: The brain integrates sensory signals at low spatial disparity (i.e.

small difference for auditory vs. visual report), but computes different spatial estimates for auditory and visual signals at large spatial disparities (Figure 5.6D).

For completeness, we also observed a significant interaction between spatial disparity and visual reliability between 45 – 100 ms and between 170 – 235 ms. This interaction resulted from a larger spatial window of integration for stimuli with low versus high visual reliability. Basically, it is easier to determine whether two signals come from different sources when the visual input is reliable.

Finally we investigated when in the time course of brain dynamics the behavioural weight indices were reflected. Hence we calculated the Pearson correlation coefficient between the behavioural and neural w_{AV} values separately for each time point. The correlation coefficient fluctuated around chance level (R = 0) until 95 ms post stimulus, then it started to rise reaching its maximum at 340 ms post stimulus (R = 0.73). This was followed by a plateau phase until 495 ms post stimulus, and then it started to slowly decay until the end of the time window. Cluster permutation test confirmed, that the correlation was significantly greater than chance, revealing a significant cluster starting at 170 ms post stimulus until the end of the time window (p < 0.001). The plateau phase with the maximum similarity between 340 – 495 ms indicates that participants' behavioural responses emerged at a later processing stage, pointing towards the involvement of higher order cortical areas.



Figure 5.6 'Model free' EEG analysis results. Audiovisual weight index as a function of visual reliability, task relevance, and disparity and its correlation with the corresponding behavioral weight index across time. Audiovisual weight index wAV (across-participants circular mean and bootstrapped 68% confidence interval; n = 13). Horizontal gray lines indicate pure auditory and visual influence (wAV value of 0° and 90° respectively). Shaded gray areas show statistical significance of effects on wAV. (A) Audiovisual weight index wAV as a function of visual reliability (high [VR+] versus low [VR-]). (B) Audiovisual weight index wAV as a function of task relevance (auditory [A] versus visual [V] report). (C) Audiovisual weight index wAV as a function of task function of audiovisual spatial disparity (small [≤ 6.6 ; D-] versus large [>6.6; D+]). (D) Audiovisual weight index wAV as a function of task relevance and disparity. As the neural wAV is extremely variable in the pre stimulus baseline period we decreased the contrast of the time courses up to 50 ms post stimulus for aesthetic reasons. (E) Circular-circular correlation (across-participants mean after Fisher z-transformation \pm SEM; n = 13) between the neural weight index wAV and the equivalent behavioral weight index in the time course.

Temporal dynamics of Bayesian Causal Inference

Finally, we investigated how estimates of Bayesian Causal Inference emerged across time in the brain. For this we trained the SVR models on spatially congruent stimuli and generalized to all trials including spatially congruent and incongruent ones as well. We used the decoded labels (identical to the ones used for the neural audiovisual weight) for further analysis. Similarly to the 'model based' behavioural analysis approach we fitted four models to the decoded labels separately for each time point: (i) 'segregation auditory', (ii) 'segregation visual' (iii) 'forced fusion' and (iv) the 'Bayesian Causal Inference' model (Figure 5.1C). We evaluated the model fits using the Bayesian Information Criterion (BIC) and performed Bayesian model selection on the group level (Rigoux et al., 2014) to compute the exceedance

probability (i.e. the probability that the model at hand is more likely than any of the others given the data). The time course of the exceedance probability revealed a sequential pattern (Figure 5.7A): Early in the time course, the 'segregation visual' model dominated up to 100 ms post stimulus. This model assumes only one cause and estimates the stimulus locations according to the visual modality only. Then the 'force fusion' model explained the neural representations best until 290 ms post stimulus, followed by the 'Bayesian Causal Inference' as the winning model after 290 ms up until 520 ms post stimulus. After this the 'segregation visual' model explained the neural representations best, however after around 500 ms post stimulus all model fits were relatively poor (Figure 5.7B), hence they must be interpreted cautiously. Collectively, these results suggest, that neural representations of audiovisual integration evolve gradually across time according to the hierarchy of the BCI model: starting with the segregation estimate, then the forced fusion and ultimately the BCI estimate.



Figure 5.7 'Model based' EEG analysis results. Time course of the exceedance probabilities (A) and corresponding log-likelihoods (B) of the four models: (i) 'segregation auditory',(segA) (ii) 'segregation visual' (segV) (iii) 'forced fusion' (Fusion) and (iv) the 'Bayesian Causal Inference' (BCI).

Discussion

We combined psychophysics, EEG, multivariate pattern decoding and computational modelling to characterize how representations of audiovisual integration emerge across time in the human brain. Our results demonstrate that multisensory interactions evolve dynamically in the brain: different computational operations emerge sequentially in line with the hierarchy of the Bayesian Causal Inference model.

Previous studies have demonstrated, that multisensory interactions are not confined to higher order association cortices, but also involve primary sensory areas via various mechanisms such as direct connectivity, thalamo-cortical interactions and top-down influences. Moreover, two recent studies by Rohe and Noppeney (Rohe & Noppeney, 2015a, 2016) have demonstrated, that audiovisual interactions emerge across a hierarchy of brain regions, governed by distinct computational principles at each level of the hierarchy. We adopted the experimental design and analysis approaches of these two studies and applied them to EEG recordings to investigate the temporal dynamics of audiovisual integration from a 'model free' (i.e. based on the relative audiovisual weight) and a 'model based' (i.e. based on models of Bayesian Causal Inference) perspective.

We found, that early processes until about 100 ms post stimulus represent audiovisual signals assuming independent causes ('full segregation'). Interestingly,

this time window was dominated by the 'segregation visual' model only and the 'segregation auditory' model was completely outperformed. We speculate that the topographical pattern of the visual evoked potential might have been modulated stronger by the spatial location than the auditory evoked response in this early time window. Similarly, the visual stimulus location dominated the neural relative audiovisual weight w_{AV} until 90 ms post stimulus and auditory stimuli influenced the representations only depending on spatial disparity exerting a repulsion effect on the w_{AV} when spatial disparity was large.

Between 100 - 290 ms post stimulus the brain represented audiovisual stimuli as if they were caused by a common cause ('forced fusion') weighting them by their relative reliabilities. Accordingly, the relative sensory reliability of the stimuli started to modulate the w_{AV} from 90 ms post stimulus, importantly the time window between 150 -220 ms was modulated solely by sensory reliability. It has been previously shown, that the ventriloquist illusion is associated with the N260 component (between 230 – 270 ms) co-localized with BOLD activity in the planum temporale (Bonath et al., 2007). Our results further corroborate this evidence, as the ventriloquist illusion is considered to be the result of reliability weighted integration of audiovisual signals (Alais & Burr, 2004). Furthermore a recent study by Boyle and colleagues (2017) investigated the temporal dynamics of auditory and visual weights in an audiovisual rate discrimination task with EEG. They also found reliability effects starting from 84 ms, further corroborating, that sensory reliability does not influence sensory processing in the earliest stages.

Finally, the veridical Bayesian Causal Inference estimate, which also best explained observers' behavioural responses, emerged after 290 ms post stimulus. Similarly, the

 w_{AV} dynamics showed, that the brain's spatial estimate incorporated both the taskrelevance and the relative sensory reliability of the stimuli from 220 – 450 ms. Importantly asignificant interaction of task-relevance and spatial disparity points towards that the integration breaks down depending on spatial disparity between 350 – 435 ms. Participants behavioural w_{AV} s correlated significantly with the neural w_{AV} s starting from 170 ms and reaching a maximal value at a plateau between 340 – 495 ms post stimulus. Critically this time window overlaps with the representation of the Bayesian Causal Inference estimate further corroborating, that the final behavioural spatial estimate emerged only after a series of distinct computations.

Collectively these results parallel the evidence on how multisensory computations emerge across a hierarchy of brain regions (Rohe & Noppeney, 2015a, 2016): First, up to 100 ms after stimulus onset the signals are segregated, then they are integrated assuming a common cause up until 290 ms, and last the veridical perceptual estimate that combines signals as predicted by Bayesian Causal Inference model is formed.
CHAPTER 6: SPATIO-TEMPORAL CHARACTERISTICS OF VISUALLY INDUCED AUDITORY SPACE ADAPTATION IN THE HUMAN BRAIN

The current chapter is based on a collaborative project between Máté Aller, Agoston Mihalik and Uta Noppeney and is being prepared for publication as a research article. MA and AM have contributed equally to the work and will be shared first-authors of the research article. The experiments were designed by MA, AM and UN. Piloting of the behavioural experiment was performed by AM. Data of the behavioural and EEG experiments were collected by MA, fMRI data were collected by MA and AM. Behavioural and fMRI analyses were performed by AM. EEG analyses, population rate code modelling, multidimensional scaling and representational similarity analyses between EEG and fMRI were performed by MA. All data analyses were supervised by UN. The manuscript was written by MA and AM supervised by UN. Part of the data (psychophysics and fMRI experiment) presented here has already been presented in the thesis of AM (Mihalik, 2017), but were included here as well due to their importance to the research question examined.

Introduction

In our ever-changing environment we are informed by multiple senses about the properties of objects and events around us. Information provided by different sensory modalities must be merged into a unified and robust percept in order to respond effectively to the challenges our environments expose us to. Our nervous system handles this task effortlessly and near-optimally despite the noisy nature of the world and our sensory system (Ernst & Bülthoff, 2004). Importantly, our sensory system is

subject to changes in its physical properties and input statistics at multiple timescales throughout our lives. For example during development, when the size and the relative distance between sensory organs is constantly changing (Gori et al., 2008, 2010); or during ageing when the acuity of vision and hearing is slowly deteriorating. The nervous system must adapt to these persistent changes to maintain the consistency between sensory modalities and to make sure, that the sensory estimates are accurate representations of the physical properties of the world.

Sensory adaptation or plasticity has been studied extensively over the past decades both on the behavioural and the neural level (Ernst & Di Luca, 2012; Ghahramani, Wolpert, & Michale, 1997; Held, 1965; Knudsen & Knudsen, 1989). One particular phenomenon, called the ventriloquist aftereffect, offers an elegant paradigm in which audiovisual recalibration can be studied (Canon, 1970, 1971; Lewald, Foltys, & Töpper, 2002; M. Radeau & Bertelson, 1974; Monique Radeau & Bertelson, 1976; G. H. Recanzone, 1998). The ventriloguist aftereffect emerges when subjects are exposed to synchronous, yet spatially discrepant auditory and visual stimuli. Exact temporal synchrony of two signals strongly implies that they represent a single event or object, but the spatial discrepancy between them is in conflict with this. The brain resolves the conflict by shifting the apparent location of the auditory signal towards the visual and vice versa. Importantly this shift persists even if the auditory or visual stimuli are presented alone. Initial evidence suggested, that recalibration occurs only after prolonged exposure to multisensory conflict (Canon, 1970, 1971; Frissen, Vroomen, de Gelder, & Bertelson, 2003; Lewald et al., 2002; M. Radeau & Bertelson, 1974; Monique Radeau & Bertelson, 1976; G. H. Recanzone, 1998; Woods & Recanzone, 2004), however these studies used designs where the adaptation trials

were blocked into prolonged segments. More recent studies used interleaved designs where the adaptation trials were interspersed between the sound localization trials. These studies demonstrated that adaptation takes place much more rapidly (Frissen, Vroomen, & De Gelder, 2012; Mendonca, Escher, van de Par, & Colonius, 2015; Wozny & Shams, 2011a), even after exposure to a single spatially discrepant audiovisual stimulus pair (Wozny & Shams, 2011b). Furthermore, it has been suggested, that the same computational principles govern multisensory recalibration as multisensory integration, namely it is proportional to the relative sensory reliabilities of the individual signals (Burge et al., 2010; Witten & Knudsen, 2005). Recent evidence however questions this and emphasizes the role of the relative accuracies of the senses (i.e. the probability with which a sensory signal represents the true magnitude of the underlying physical property) (A. Zaidel et al., 2011; Adam Zaidel et al., 2013).

Despite the numerous studies characterizing the ventriloquist aftereffect at the behavioural level, the neural substrates remain unknown. Several key brain areas have been hypothesised to be involved in audiovisual spatial recalibration such as primary auditory cortex, planum temporale, parietal cortices, frontal eye fields and superior colliculus. These were proposed based on behavioural (e.g. frequency and spatial selectivity, spatial reference frames) and neurophysiological (mainly from experiments on barn owls (McAlpine, 2005)) characteristics of the process. For example, early studies proposed the primary auditory cortex as a candidate area as initial evidence suggested, that the ventriloquist aftereffect was frequency specific (Lewald, 2002; G. H. Recanzone, 1998). However, multiple behavioural studies have since demonstrated the transfer of the ventriloquist aftereffect across wide frequency

ranges (Frissen et al., 2004, 2003) considerably extending the pool of possible candidates regions.

The present study combined psychophysics, functional magnetic resonance imaging (fMRI), electroencephalography (EEG) and advanced multivariate decoding models to investigate the spatial (fMRI) and temporal (EEG) characteristics of visually induced auditory adaptation in the human neocortex. Participants were presented with unisensory auditory signals before and after adaptation periods of spatially conflicting audiovisual signals. We trained support vector regression models on the BOLD and EEG response patterns elicited by unisensory auditory stimuli before adaptation separately for five pre-defined regions of interest along the auditory spatial processing hierarchy (fMRI) and each time point (EEG). Then we applied the learnt mapping to decode auditory spatial locations before and after adaptation and used them as a basis of a series of analyses to characterize the coding and recalibration of auditory space. First, we assessed the extent to which each brain region and time point coded auditory space and recalibration. Then we applied representational similarity analysis (Kriegeskorte, 2008) to elucidate the representational details of space and recalibration coding and to compare representations between imaging modalities and model predictions based on population rate code model (Salminen et al., 2009). Furthermore, we fitted psychometric functions to participants' responded locations and neurometric functions to the decoded locations to characterize the remapping of auditory space both at the behavioural and neural level.

Methods

Behavioural experiment

Participants

Fifteen right-handed participants (10 females, mean age = 22.1; SD = 4.1) were selected from a pool of nineteen volunteers to take part in the behavioural experiment. Participants were selected based on the following criteria: (i) no history of neurological or psychiatric illness; (ii) normal or corrected-to-normal vision; (iii) reported normal hearing; (iv) accurate sound localization abilities; and (v) high accuracy in the adaptation task. All the participants gave informed consent to participate in the experiment and received monetary compensation. Participants attended a pre-screening day before the main experiment to determine their performance in sound localization and adaptation tasks (for a detailed description, see the 'Pre-screening' section). The study was approved by the research ethics committee of the University of Birmingham and was conducted in accordance with the principles outlined in the Declaration of Helsinki.

Experimental procedure

Participants were screened in a separate day before the main experiment in sound localization as well as adaptation tasks of the main experiment. Participants selected for the experiment attended 4 days: 2 days with the V stimulus on the left during adaptation (VA-posttest) and 2 days with the V stimulus on the right during adaptation (AV-posttest). The direction of the adaptation was kept constant within a day and was systematically changed across days, counterbalanced between

participants. The responding hand for catch trials was altered between runs, also counterbalanced between participants.

Each day consisted of 5 pretest periods followed by 10 adaptation periods interleaved with 10 posttest periods. A typical day lasted for ~1.5 hours and was divided up to 3 parts with breaks in between. Each part consisted of 5 test periods (with an adaptation period preceding each posttest period), and was completed in one sitting. The forehead of participants was marked at the beginning to ensure the same head positioning throughout the day and avoid any variability in sound localization within a day due to head positioning. Participants were instructed to fixate to a central fixation cross throughout the experiment.



Figure 6.1 Overall layout, design, task and stimuli of the experiment. (A) Overall layout of the experiment. (B) The experimental design is divided into 3 phases: pretest, adaptation and posttest. Most trials are with no response. Response trials are indicated by a dimming fixation cross (500 ms after A onset) in the sound localization task (during pre- and posttest) and by a dimming V stimulus (at AV onset) in the visual localization task (during adaptation). Abbreviations: A, auditory; AV, audiovisual; RDM, representational dissimilarity matrix; SOA, stimulus onset asynchrony.

Stimuli

The visual (V) stimuli used in the adaptation task were clouds of 15 white dots (diameter = 0.4° visual angle) sampled from a bivariate Gaussian presented on a dark grey background (90% contrast) for 50 ms in synchrony with the auditory stimuli. The horizontal and the vertical standard deviation of the Gaussian were set to 1.5° visual angle. The auditory (A) stimuli, used in all the tasks, were 50 ms long bursts of white-noise with 5ms on/off ramp. The sound stimuli were filtered through a generic HRTF of the MIT database using normal pinnae (Gardner & Martin, 1995, http://sound.media.mit.edu/resources/KEMAR.html). Measurements of the MIT database were interpolated to generate sound stimuli originating from the desired spatial locations. Scanner background noise was superimposed on the sound stimuli to match the task environment in the fMRI experiment.

Pre-screening

To familiarize participants with the sound localization task, they performed a 1 minute long practice run. Overall, participants performed 3 tasks during the pre-screening: (1) sound localization task of A stimuli responding on all trials; (2) sound localization task of A stimuli responding on catch trials only; (3) adaptation task responding on catch trials only. Task (2) and (3) were identical to the tasks of the main experiment. Task (1) was similar to task (2) with the exception of participants responding on all trials and the presentation of stimuli not being optimized (no pseudo-randomization, no fixation periods). Participant selection for the main experiment was based on the following criteria:

- JND <4° in task (1),
- JND <4°and d'>2.5 in task (2),

• d'>2.5 in task (3).

For the calculation of the indices, see the Behavioural data analysis section.

Experimental design

The experiment was divided into 3 phases: pre-adaptation test, adaptation and postadaptation test. The pre- and post-adaptation phases are called pre- and posttests or test phases, collectively. In the test phases, participants performed a sound localization task on A stimuli indicating a response only on catch trials. In the adaptation phase a visual detection task was performed. Throughout the entire experiment participants were instructed to fixate a central cross (0.5° diameter).

In the test phases, participants were presented with A stimuli with stimulus onset asynchrony (SOA) = 2 s with \pm 0.2 s jitter (sampled from a uniform distribution). The stimuli were sampled from spatial locations $\pm 12^{\circ}$, $\pm 5^{\circ}$, $\pm 2^{\circ}$ and 0° visual angle. To match the task with the follow-up fMRI study, stimuli were presented in blocks (~28 sec) separated by fixation periods (6 sec). In addition, stimuli were pseudorandomized using repetitions of 4, 3, 2 and 1 stimuli improving the design efficiency (again, aimed for the fMRI study). One period of stimulation consisted of 90 stimuli (separated into 5 stimulus blocks and 4 fixation periods) and lasted for ~3.5 min. Participants localized the A stimulus in a two alternative (left-right) forced choice fashion and were instructed to respond only on catch trials (~22%). Catch trials were indicated by a cue 500 ms after stimulus onset, which consisted of a 55% contrast change in the fixation cross lasting for 200 ms. To indicate the A location participants pressed either of two buttons with their index and middle fingers respectively, alternating their responding hand between the three parts of the day. The order of the responding hand was counter-balanced across days to avoid any motor response confound.

In the adaptation phase, participants were presented with audiovisual stimuli with SOA = 0.5 s (without jitter). In the first adaptation period (after break) 360 audiovisual stimuli (3 min) were presented, whereas 120 audiovisual stimuli (1 min) were presented in the remaining adaptation periods. On each trial the V stimulus emerged from three possible spatial locations (-5°, 0°, 5°) along the horizontal midline of the screen. The A stimulus was spatially shifted 15° to the right (for VA adaptation) or left (for AV adaptation) with respect to the V stimulus. The spatial locations of the audiovisual stimuli were kept identical within sets of five consecutive presentations and it was changed randomly in between. To maintain participants' attention they were instructed to detect an occasional 20% contrast change of the V stimuli which occurred on 10% of the trials (catch trials).

Experimental setup

Visual and auditory stimuli were presented using Psychtoolbox version 3.0.11 (Brainard, 1997; Kleiner et al., 2007) running under MATLAB R2011b (MathWorks Inc.) on a MacBook Pro (Mac OSX 10.6.8). Participants were seated at a distance of 600 mm from a 24" LCD screen (iiyama ProLite B2483HS) resting their head on a chin rest. Two accessory rods were mounted on the chin rest serving as forehead rest and allowing stable head positioning. Auditory stimuli were delivered via circumaural headphones (Sennheiser HD 280 Pro). Participants used a standard USB keyboard for responding.

Eye movement recording and analysis

To address potential concerns that results may be confounded by eye movements, we evaluated the eye movements of the participants. Eye recordings were calibrated in the recommended field of view (32° horizontally and 24° vertically) for the EyeLink 1000 Plus system with the desktop mount at a sampling rate of 2000 Hz. Eye position data were on-line parsed into events (saccade, fixation, eye blink) using the EyeLink 1000 Plus software. The 'cognitive configuration' was used for of saccade detection (velocity threshold = 30°/sec, acceleration threshold = 8000°/sec², motion threshold = 0.15°) with an additional criterion of radial amplitude larger than 1°. Events occurring from 0 to 500 ms post-stimulus were selected and further analysed. For the eye movement results of the psychophysics experiment, see the next section. In the EEG experiment, individual trials were rejected based on eye movement results (see EEG Preprocessing section), but no further analysis was performed. In the fMRI experiment, precise positioning of the participant's head inside the scanner bore was absolutely critical for the sensitive measurement of spatial recalibration, which did not allow us high quality eye movement recordings.

Eye movement results for psychophysics experiments

Fixation was well maintained throughout the entire experiment with participants fixating correctly in 94.02% \pm 1.00% (mean \pm SEM) of all trials (i.e. eye gaze was maintained within a radius of 2° and without saccades or eyeblinks) and post-stimulus saccades detected in only 1.24% \pm 0.33% (mean \pm SEM) of all trials (including response and non-response trials in all test and adaptation phases). To exclude the possibility that eye movement related processes confounded the results of our main analysis we tested if there was a difference in eye movement indices

across the 3 conditions of the experimental test phases (A pre-test, AV and VAposttest). For this we focused on non-response trials only. Eye movement indices of % saccades, % eye blinks, and post-stimulus mean horizontal fixation position were quantified for each of the 3 test phase conditions and entered into one-way repeated measures ANOVAs.

No significant differences were observed across the 3 test phase conditions for % saccades and % blinks. A significant effect was found on the horizontal mean eye position (F(2,14)=8.02, p=0.002). Fisher's LSD post-hoc tests revealed differences between pretest and AV-posttest, and pretest and VA-posttest phases (p=0.004 and p=0.027 for VA and AV-posttest respectively). Importantly, the mean horizontal fixation position did not vary systematically with adaptation direction (mean \pm STD: - 0.38° \pm 0.07° and -0.52° \pm 0.09° for AV- and VA-posttest respectively) and the direct comparison between AV and VA-posttest was not significant.

Behavioural data analysis

Signal detection measures

Hit rates and signal sensitivity measure d-prime was calculated on catch trials for both sound localization and adaptation tasks as follows: $d' = Z(\text{probability}_{hits}) - Z(\text{probability}_{false alarms})$. The calculation assumed equal-variance Gaussian for both the signal and noise distributions. 100% hit rate and 0% false alarm rate were considered as 99.999% and 0.001%, respectively, to enable the calculation of z-values (otherwise infinite).

Psychometric function fitting

For the pre-, VA and AV post-adaptation tests we estimated psychometric functions to the unisensory auditory spatial discrimination responses that we obtained from the 22% 'response trials'. We fitted cumulative Gaussian psychometric functions (PF) to the probability of 'perceived right responses' as a function of stimulus location (±12°, $\pm 5^{\circ}$, $\pm 2^{\circ}$ and 0°) (Figure 6.4) individually for each participant, using maximumlikelihood estimation as implemented in the Palamedes toolbox (<u>www.palamedestoolbox.org</u>). To enable reliable parameter estimates for each participant, we employed a multi-condition fitting using the following constraints: i) the just noticeable differences (JND, i.e. perceptual threshold or slope parameter) were set equal across all conditions; ii) guess and lapse rates were set equal to each other and equal across all conditions. Further, we constrained the fitted guess and lapse rate parameters to be within 0 and 0.1.

Statistical inference of PFs was made in 3 steps. At first, we tested the assumption that the data for each condition and participant can be accurately fitted by a cumulative Gaussian function. To validate this assumption, we performed a goodness of fit test via likelihood ratio test. This test compares i. the likelihood of participants' responses given the model that is constrained by cumulative Gaussian function (i.e. our 'target model') to ii. the likelihood given a so-called 'saturated' model that models observers' responses with one parameter for each stimulus location in each condition. The resulting likelihood ratio for the original data set is then compared with a null-distribution of likelihood ratios generated by parametrically bootstrapping the data (5000x) from the 'target model' and refitting the 'target' and 'saturated models' (Kingdom & Prins, 2016; Wichmann & Hill, 2001). As in each of

the participants, the likelihood ratio for the original data set was greater than 5% of the parametrically bootstrapped likelihood ratios (i.e. p > 0.05), we inferred sufficient goodness of fit for all participants.

Secondly, we performed Bayesian model comparison at the level of participants to assess whether AV/VA adaptation induced a shift in participants' perceived location of unisensory A stimuli. More specifically, we compared i. a 'static or no adaptation' model' with an 'adaptation model'. The 'staticmodel' assumes no effect of audiovisual spatial adaptation on unisensory auditory spatial representations by constraining the point of subjective equality (PSE) for the pre-adaptation, post-VA-adaptation and post-AV-adaptation PFs to be equal. The 'adaptation model' includes three PSE values for the pre-adaptation, post-VA-adaptation and post-AV-adaptation PFs thereby accommodating potential shifts in the PSF as a result of audiovisual recalibration. For each participant, we calculated log-likelihood for each of the two models and computed the Akaike Information Criterion (AIC, Akaike, 1974) as an approximation to the model evidence (Rigoux et al., 2014). As in each of the participants, the likelihood ratio for the original data set was less than 5% of the parametrically bootstrapped likelihood ratios (i.e. p < 0.05), and the AIC of the 'static model' was higher than the AIC of the 'adaptation model' we inferred that the 'adaptation model' better explained participants' data.

Finally, we performed Bayesian model comparison at the group level to allow population level inference (Rigoux et al., 2014). We used model comparison as it is implemented in SPM12 (Friston, Jezzard, & Turner, 1994) based on the expected posterior probability (the probability that a given model generated the data for a randomly selected participant), the exceedance probability (the probability that a

given model is more likely than any other model), and the protected exceedance probability (additionally accounting for the fact that differences in model frequencies may be due to chance).

fMRI experiment

Participants

Six participants (4 females, mean age = 22.2; SD = 3.7) were selected from the participants of the behavioural experiment, who showed the largest recalibration effects (for details, see respective section). One of the participants was an author of the study (AM). One participant was excluded after 3 days due to poor detection performance in the adaptation phases indicated by an overall d' < 2.5 (the same criterion was used for pre-screening participants in the behavioural experiment). All participants gave informed consent to participate in the fMRI experiment and received monetary compensation. The study was approved by the human research ethics committee at the University of Birmingham and was conducted in accordance with the principles outlined in the Declaration of Helsinki.

Experimental procedure

Participants performed 4 days similarly to the behavioural experiment. Here, the direction of the adaptation was kept constant within a day and was systematically changed across days, counterbalanced between and within participants. Each day consisted of 10 pretest periods followed by 8 adaptation periods interleaved with 8 posttest periods. In addition, periods were organized into runs and the scanner was restarted between each run. One run was composed of 2 pretest periods or 2 adaptation periods and 2 posttest periods. The responding hand for catch trials was altered between runs, also counterbalanced between participants. Participants

performed a short left-right discrimination task in the beginning of each day to check that localization is unbiased in the scanner environment. Participants were instructed to fixate on a central fixation cross throughout the experiment.

Stimuli

Stimuli used in the fMRI experiment were identical to the behavioural experiment except for the background scanner noise, which in this case was not artificially provided by the experimenter, but by the scanner itself.

Experimental design

The experimental design was identical to the behavioural experiment except for the following details: (i) 360 audiovisual stimuli (3 min) were presented in all adaptation periods (except for the first 2 participants, where 1 min adaptations were still used). This change enabled a more robust adaptation effect. (ii) The spatial location of the audiovisual stimuli was kept identical within sets of 20 consecutive presentations and it was changed randomly in between. This change was made to improve design efficiency for the audiovisual signals. (iii) 10 s long fixation intervals were introduced at the beginning and end of runs as well as between periods to improve estimation of the baseline fMRI signal.

Experimental setup

The experimental setup was similar to the behavioural experiment except for the followings: (i) the visual stimuli were back projected to a plexiglas screen using a D-ILA projector (JVC DLA-SX21) and were made visible to the subject through a mirror mounted on the magnetic resonance (MR) head coil. (ii) Auditory stimuli were delivered via a pair of MR compatible headphones (MR Confon HP-VS03). We

showed in the preceding behavioural study that eye movements did not confound the aftereffect, and decided not to record eye data in the scanner.

MRI data acquisition

We used a 3T Philips Achieva scanner to acquire both T1-weighted anatomical images (TR/TE/TI, 7.4/3.5/min. 989 ms; 176 slices; image matrix, 256 x 256; spatial resolution, 1 x 1 x 1 mm³ voxels) and T2*-weighted echo-planar images (EPI) with blood oxygenation level-dependent (BOLD) contrast (fast field echo; TR/TE, 2800/40 ms; 38 axial slices acquired in ascending direction; image matrix, 76 x 75; slice thickness, 2.5 mm; interslice gap, 0.5mm; spatial resolution, 3 x 3 x 3 mm³ voxels). There were 20 pretest runs, each with 160 volumes over 4 days. There were 16 posttest runs both for VA and AV-posttest over 2 days, respectively. The first 4 volumes were not acquired to allow T1 equilibration effects.

Behavioural data analysis

The behavioural results obtained in the scanner were analysed similar as in the preceding behavioural study.

fMRI analysis

Pre-processing

The data were analysed with Statistical Parametric Mapping (SPM12; Wellcome Trust Centre for Neuroimaging, London, UK; http://www.fil.ion.ucl.ac.uk/spm/; Friston et al., 1995; Friston, Holmes, Worsley, Frith, & Frackowiak, 1995; Friston, Jezzard, & Turner, 1994). Scans from each participant were realigned using the first as a reference, unwarped and corrected for slice timing. The time series in each voxel was high-pass filtered to 1/128 Hz. The EPI images were spatially smoothed with a

Gaussian kernel of 3 mm FWHM and the data were analysed in native participant space. The data were modelled in a mixed event-block related fashion with regressors entered into the design matrix after convolving the unit impulse (representing a single trial) or the block with a canonical hemodynamic response function and its first temporal derivative. Unisensory sound location conditions (of the test phases) were modelled as events and visual location conditions (of the adaptation phase) as blocks. Moreover, sound locations for trials with and without responses were modelled separately, as well as all responses within visual blocks were modelled in a separate event-related regressor. Realignment parameters were included as nuisance covariates to account for residual motion artefacts. Conditionspecific effects for each subject were estimated according to the general linear model (GLM) separately for each run. We applied multivariate noise normalization to the beta estimates via the noise covariance matrix obtained from the residuals of the GLM (Walther et al., 2016) and using the optimal shrinkage method described in (Ledoit & Wolf, 2004). As a last pre-processing step, each beta image was normalized dividing by its Euclidean norm. Due to the sluggish nature of BOLD response, only non-catch trials were included in all subsequent fMRI analyses to avoid response related changes in BOLD activations during catch trials.

Regions of interest definition

We used five auditory regions of interest (ROI) based on anatomical and neurophysiological studies in non-human primates as well as human neuroimaging studies characterizing auditory spatial processing (Arnott, Binns, Grady, & Alain, 2004; Bushara et al., 1999; Kong et al., 2014; Michalka, Rosen, Kong, Shinn-Cunningham, & Somers, 2016; Rauschecker & Tian, 2000; Weeks et al., 1999). All

regions were combined from the left and right hemispheres, unless otherwise stated. HG (Heschl's gyrus), hA (higher auditory cortex, including planum temporale) and IPL (inferior parietal lobule) were defined using the following parcellations of the Destrieux atlas of Freesurfer 5.3.0 (Dale, Fischl, & Sereno, 1999; Destrieux, Fischl, Dale, & Halgren, 2010): i) anterior transverse temporal gyrus for HG; ii) transverse temporal sulcus, planum temporale and posterior ramus of the lateral sulcus for hA; iii) supramarginal gyrus and inferior part of the postcentral sulcus for IPL. IPS (intraparietal sulcus) and FEF (frontal eye-field) were defined using the following group-level retinotopic probabilistic maps (Wang, Mruczek, Arcaro, & Kastner, 2015): i) IPS0, IPS1, IPS2, IPS3, IPS4, IPS5 and SPL1 for IPS; ii) hFEF for FEF. All probabilistic maps were thresholded at 10 percentiles and inverse normalized to subject space.

Multivariate decoding

We trained linear support vector regression (SVR) model in a 4-fold stratified crossvalidation scheme as implemented in LIBSVM 3.17 (Chang & Lin, 2011) to accommodate the continuous nature of the auditory locations. More specifically, the voxel response patterns were extracted in a particular region of interest (e.g. A1) from the pre-processed beta image estimates corresponding to the BOLD response for each auditory location condition and run of the GLM as discussed above. SVR models were trained to learn the mapping from the condition-specific fMRI responses patterns (i.e., examples) to the condition specific spatial locations (i.e., labels) using examples from the pretest conditions from all but one fold. Then the learnt mapping was used to decode the spatial locations from the voxel response patterns of both the pretest and posttest examples in the remaining fold. This procedure was

repeated for all folds, yielding one predicted spatial location for each example of the pretest runs. Default SVR hyper-parameters (C = 1, v = 0.5) were used to train the models.

Decoding accuracy was evaluated as the Fisher transformed Pearson correlation coefficient between the true and decoded auditory locations. To assess whether the decoding accuracy was better than chance we entered the Fisher transformed Pearson correlation coefficients into a between subjects bootstrap-based t-test against zero (Efron & Tibshirani, 1993).

The recalibration index (RI) was computed as follows: first the predicted auditory location was binarized (left vs. right) and the proportion of 'decoded right responses' was computed over VA- or AV-posttest examples. We then subtracted the proportion of VA-posttest from the proportion of AV-posttest. We used the same statistical method for significance testing as for the decoding accuracy.

 $RI = p(decoded right)_{AV} - p(decoded right)_{VA}$

Neurometric function fitting

We used the predicted spatial locations from multivariate decoding to plot the percentage 'decoded right responses' as a function of stimulus location $(\pm 12^\circ, \pm 5^\circ, \pm 2^\circ \text{ and } 0^\circ)$. Similarly to the behavioural analysis, we fitted cumulative Gaussian functions to all pre- and posttest data using maximum-likelihood estimation as implemented in Palamedes toolbox (www.palamedestoolbox.org). Due to the lower signal-to-noise ratio of fMRI data, NFs were fitted on data averaged across all participants. To enable reliable parameter estimates, we employed a multi-condition

fitting using the following constraints: i) the just noticeable differences (JND, i.e. perceptual threshold or slope parameter) were constrained to be equal across all conditions; ii) guess and lapse rates were set to be equal and constrained to be equal across all conditions; iii) lapse rates were set to be within 0 and 0.45. This method enabled a similar approach to the behavioural analysis, but now at the neural level.

For statistical inference, we used model comparison via likelihood-ratio test. At first, we performed a goodness-of-fit test using a 'saturated model' and a 'target model' as described in the behavioural analysis. Briefly, the 'saturated model' assumes one parameter for each auditory location in each test phase (pretest, VA and AV-posttest). The 'target model' is our 'full model' (3 NF model) assuming five parameter estimates: 3 PSEs, 1 JND and 1 guess/lapse rate. We calculated the likelihood ratio for the original data and compared it with a null-distribution of likelihood ratios generated by parametrically bootstrapping data from the 'saturated' and 'target' models (Kingdom & Prins, 2010). Significance was tested at $\alpha = 0.05$ level.

Secondly, we tested the effect of adaptation using a 'null model' and a 'full model' as described in the behavioural analysis and above. Briefly, the 'null model' assumes three parameter estimates: 1 PSE, 1 JND and 1 guess/lapse rate; the 'full model' assumes five parameter estimates: 3 PSEs, 1 JND and 1 guess/lapse rate. The likelihood ratio for the original data was then compared with a null-distribution of likelihood ratios described above.

Representational similarity analysis

At first, mean beta images were calculated for the 7 pretest conditions (space coding) and the 7 VA-posttest + 7 AV-posttest conditions (recalibration coding) across runs. Then RDMs were computed separately for space coding (7x7 matrix; Figure 6.2B left) and recalibration coding (14x14 matrix; Figure 6.2B right). We used the Mahalanobis distance as the dissimilarity measure (technically the squared Euclidean distance was calculated, but this is equivalent to the Mahalanobis distance since the data were pre-whitened, see eq. 7 in (Walther et al., 2016)). RDMs were rank transformed and scaled between 0 and 1. For the recalibration coding RDMs the four 7 x 7 submatrices were individually rank transformed and scaled.

To assess the relatedness between the RDMs from fMRI, EEG and the models we computed the Spearman product moment coefficient between the lower triangular matrices of the respective RDMs for each participant separately (technically the Pearson correlation coefficient was computed but this is equivalent to the Spearman correlation coefficient as the RDMs were rank transformed previously). Then the correlation coefficients were Fisher z-transformed and averaged over participants. For significance testing the correlation coefficients were Fisher transformed and the same statistical method was used as for the decoding accuracy.

EEG experiment

Participants

Five participants who previously took part in the fMRI experiment participated in the EEG experiment (4 female, mean age = 22.2 years, SD = 3.7 years). All participants gave their informed consent to the experiment. The study was approved by the

research ethics committee of the University of Birmingham and was conducted in accordance with the principles outlined in the Declaration of Helsinki.

Experimental procedure

Similarly to the fMRI experiment participants performed 4 experimental days (except for one participant, who performed two days of double length because of logistical reasons). Each day included 10 pretest periods and 8 adaptation periods interleaved with 8 posttest periods. To mimic the experimental circumstances in the fMRI environment as closely as possible periods were organized into runs as in the fMRI experiment. A short sound localization task was performed at the beginning of each day to check whether the in-ear earphones were symmetrically fitted. The direction of the adaptation was kept constant within a day and was systematically changed across days, counterbalanced between participants. The responding hand for catch trials was altered between runs, also counterbalanced between participants.

Stimuli and experimental design

We used the same stimuli as in the behavioural and fMRI experiments. The same fMRI scanner background noise was played throughout the experimental runs. The experimental design was identical to the fMRI experiment except the 6 second fixation events and the 10 second fixation periods at the beginning and the end of the runs were omitted.

Experimental setup

The experimental setup was identical to the setup in the behavioural experiment except the device for delivering auditory stimuli. To avoid EEG artefacts generated by the electromagnetic field of the headphones we used in ear earphones (E-A-RTONE

GOLD, 3M Company Auditory Systems, Indianapolis, IN, USA), which places the electrical components far enough from the EEG electrodes. The proper fitting of the device was checked by a short sound localization task at the beginning of each day.

Eye movement recording and analysis

As in the behavioural experiment we continuously monitored the eye movements to avoid confounds. The setup and the settings for eye tracking were identical to the one in the behavioural experiment.

EEG data acquisition

Continuous EEG signals were recorded from 64 channels using Ag/AgCI active electrodes arranged in 10-20 layout (ActiCap, Brain Products GmbH, Gilching, Germany) at a sampling rate of 1000 Hz.

Behavioural data analysis

We performed the analysis on the behavioural data from the EEG experiment as described in the behavioural experiment.

EEG data analysis

Pre-processing

Channel impedances were kept below $10k\Omega$. Pre-processing was performed with the FieldTrip toolbox (Oostenveld et al., 2011). Raw data were high pass filtered at 0.1 Hz, re-referenced to average reference, and low pass filtered at 45 Hz. Bad channels were rejected based on visual inspection prior to re-referencing and interpolated using the neighbouring channels. Trials were extracted with 100 ms pre-stimulus and 500 ms post-stimulus period, baseline corrected and down-sampled to 200 Hz. Trials containing artefacts were rejected based on visual inspection visual inspection. Furthermore, trials

were rejected if they (i) contained eye blinks or (ii) saccades or if (iii) the eye gaze was away from the fixation cross further than 2 degrees. We applied multivariate noise normalization to the individual trials, estimating the noise covariance matrix separately for each time point (Guggenmos, Sterzer, & Cichy, 2017) and using the optimal shrinkage method described in (Ledoit & Wolf, 2004). Furthermore, the topographies were divided by their Euclidean norm separately for each time point and trial. Since catch and non-catch trials were identical up until 500 ms post stimulus we also included the catch trials in all subsequent analyses.

Multivariate decoding

We trained linear support vector regression models (SVR, LIBSVM (Chang & Lin, 2011)) separately for each time point up to 500 ms post stimulus. SVR models were trained in a 4-fold stratified cross-validation procedure. First, data from all conditions were randomly split into four folds making sure all spatial locations from all conditions are equally well represented in each fold. The SVR model was trained to learn the mapping from activation patterns to auditory space using examples from the pretest conditions from all but one fold. Then the learnt mapping was used to decode the spatial location from the evoked potential patterns of the pretest and posttest examples of the remaining fold. This procedure was repeated for all folds. The hyper parameters of the SVRs were fixed to their default values as defined in LIBSVM (C = 1, v = 0.5). To maximize the signal to noise ratio evoked potentials were first computed by averaging over sets of 16 randomly assigned individual trials from the same condition (Grootswagers et al., 2017). SVR models were trained on data from 50 ms time windows which were shifted across the extent of the evoked potentials from the stimulus in increments of 5 ms (Grootswagers et al.,

2017; Ramkumar, Jas, Pannasch, Hari, & Parkkonen, 2013). To avoid shifting the time courses forward in time, the onset of the windows was registered on their last time point. To increase the reliability of the estimates the whole procedure was repeated fifty times with new random grouping of the trials for the evoked potentials and the results were averaged.

Decoding accuracy was evaluated as the Fisher transformed Pearson correlation coefficient between the true and decoded auditory locations. To assess whether the decoding accuracy was better than chance anywhere over the time course we entered the Fisher transformed Pearson correlation coefficients into a between subjects bootstrap-based t-test against zero (Johnson, 2001). To control the false positive rate for multiple comparisons a cluster size based correction method was used (Nichols & Holmes, 2003).

The recalibration index (RI) was computed for each time point identically to the fMRI decoding.

Representational similarity analysis

First, evoked potentials were computed for the pretest and posttest conditions by averaging all trials within each condition. The RDMs were subsequently computed identically to the fMRI experiment for each time point separately. To match the decoding results RDMs were computed using pooled data from a 50 ms time window which was shifted across the extent of the evoked potentials in increments of 5 ms. RDMs were rank transformed and scaled between 0 and 1. For the recalibration coding RDMs the four 7 x 7 submatrices were individually rank transformed and scaled.

Multidimensional scaling

For space and recalibration coding the respective group average (first order) RDMs were projected into a one-dimensional space using non-classical multidimensional scaling (MDS) with non-metric scaling. We have chosen a one dimensional (horizontal) projection as it intuitively coincides with the horizontal space which the RDMs represent. Individual objects were colour coded according to their represented spatial location. For recalibration coding the two sets of seven conditions (VA-posttest respectively) were vertically offset for illustrational reasons, vertical distance does not represent similarity here.

For the second order MDS we computed the dissimilarity pairwise between the group level average RDMs for ROIs (fMRI), time windows (EEG) and models as their Pearson's correlation subtracted from one (1-R). The resulting second order dissimilarity matrix was then projected into a two-dimensional space.

Bootstrap tests

Throughout the study we used a non-parametric between subjects bootstrap hypothesis test based on the one-sample t-statistic (see chapter 16.4 in (Efron & Tibshirani, 1993) for a detailed description). Briefly, the one sample t-statistic was computed for the mean of the observed sample (Fisher transformed correlation coefficients, recalibration index). Then the between subject mean was subtracted from each value in the sample translating the observed sample into a sample which has a mean of zero (the null hypothesis). Then a null distribution of t-values was generated by resampling from the translated sample with replacement and computing the one-sample t-statistic for each bootstrap sample. A right-tailed p-value

was computed as the proportion of t-values in the bootstrapped t-distribution greater than the observed t-value.

Modelling

There has been a long-standing and ongoing debate whether auditory space is coded using a topographical place code or a population rate code in the neocortex. More recent evidence including neuroimaging studies from humans (Salminen et al., 2009; Trapeau & Schönwiesner, 2015) and nonhuman primates (Ortiz-Rios et al., 2017) suggests, that the latter is more likely to be used hence we modelled the neural responses using the population rate code model described by Salminen et al.(2009). Briefly, 100000 neurons were simulated for each hemisphere with Gaussian tuning curves with constant standard deviation of 67°. The means of the tuning curves were sampled uniformly from 80° to 100° or from -80° to -100° azimuth, creating an ipsilaterally and a contralaterally tuned population of neurons for each hemisphere. The ratio of the ipsi- and contralaterally tuned neurons was 30%/70% respectively for both hemispheres. Furthermore, the stimulus locations could be shifted at the input of the model to accommodate the shift from spatial recalibration. For the space coding model we simulated neural responses from the seven experimental stimulus locations. For the recalibration model we simulated neural responses from the same seven locations twice, shifting the input locations first to the left and then to the right. The value of the shift parameter was determined using the results from the behavioural experiment: it was half of the difference between the PSE value in VA and AV-posttest conditions respectively.

To compare the representations predicted by the model to the ones observed in fMRI and EEG we computed the RDMs between the simulated neuronal responses the

same way as for fMRI and EEG (see respective sections and Figure 6.2B-C). The squared Euclidean distance was used as a dissimilarity measure.



Figure 6.2 Population rate code model. (A) Population rate code model neural response functions to spatial stimulation (inset shows the entire horizontal space) (B) General layout of the representational dissimilarity matrices (RDMs) for space (left) and recalibration coding (right). (C) Space model and recalibration model RDMs predicted by the population rate code model (top) and their projections into one dimension using multidimensional scaling (MDS) (bottom). The horizontal and vertical solid white lines in the recalibration coding RDM separate the within- and between-posttest subparts, and the dashed lines indicate the diagonal of the between-posttest subpart. For recalibration coding MDS plot (bottom right) the two sets of seven conditions (VA-posttest and AV-posttest respectively) were vertically offset for clarity.

Results

To characterize how the human brain encodes and recalibrates auditory space we combined psychophysics, fMRI and EEG with multivariate pattern decoding,

representational analysis (with neurometric functions and representational similarity analysis) and computational modelling. Figure 6.3 shows the general neuroimaging analysis pipeline. Participants performed the same experimental task in a series of three experiments: first a purely behavioural, second during fMRI scanning and third during EEG recording. Each experiment spanned over four days (Figure 6.1A). The experimental design consisted of pretest, adaptation and posttest phases to assess visually induced auditory adaptation. During pre- and posttest phases, participants were presented with auditory (A) stimuli originating from 7 spatial locations (±12°, $\pm 5^{\circ}$, $\pm 2^{\circ}$ and 0°) along the horizontal plane and responded on catch trials in a forcedchoice left-right discrimination task (Figure 6.1B). During the adaptation phase, participants were presented with spatially discrepant audiovisual stimuli and attended to the visual (V) stimuli performing a detection task (Figure 6.1B). We expected that after adaptation – in the posttest phase – participants perceived the A stimuli towards the previously displaced V stimuli indicating a recalibration of the auditory system. Importantly, participants performed adaptation periods towards both directions (V stimulus to the left or right with respect to the A stimulus), which enabled us to compare recalibration effects both on the behavioural and neural levels without temporal confound.



Figure 6.3 Neuroimaging analysis pipeline. A General outline of how fMRI and EEG activity patterns were used for decoding and characterization. B Outline of decoding analysys resolved in space (fMRI) and time (EEG). C Outline of representational similarity analysis for the characterization of auditory space coding and recalibration.

Behavioural results

We fitted psychometric functions (PF) to the percentage 'perceived right responses' as a function of stimulus location for pre- and posttest behavioural responses. We hypothesised that VA-posttest would lead to a rightward shift of the function (due to fewer right responses) and AV-posttest would lead to leftward shift of the function (due to more right responses). Figure 6.4 illustrates the group level PF fits separately for the three experiments (psychophysics, fMRI, EEG). In accordance with our hypothesis, Bayesian model comparison revealed that the full model allowing for adaptation effects, i.e. changes in PSE values between pretest, VA- and AV-posttest is much more likely than the null model not allowing for any changes in PSE values (psychophysics: protected exceedance probability = 1; fMRI: protected exceedance probability = 0.888; EEG: protected exceedance probability = 0.888; Table 6.1). Table 6.2 shows the group level PSE values estimated by the full model.



Figure 6.4 Behavioural results. (A) Psychometric function (PF) fits to the behavioural data of the three experiments (behavioural, fMRI and EEG respectively). PFs are fitted separately for pretest, VA-posttest and AV-posttest conditions (B) Bar graphs represent the respective group mean PSE values. Error bars show the SEM. Individual PSE values are overlaid as scatter plots.

		Null model	Full model
	number of parameters	3	5
Psychophysics	Relative AIC	0	630.7
	Exp. post. prob.	0.059	0.941
	Exceedance prob.	0	1.000
	Prot. exceedance prob.	0	1.000
fMRI	Relative AIC	0	386.0
	Exp. post. prob.	0.11	0.890
	Exceedance prob.	0	1.000
	Prot. exceedance prob.	0.112	0.888
EEG	Relative AIC	0	328.4
	Exp. post. prob.	0.11	0.890
	Exceedance prob.	0	1.000
	Prot. exceedance prob.	0.112	0.888

Table 6.1 Behavioural results of the Bayesian model comparison between the null model and the full model. Null model (1 PF model): one underlying PF resulting in three parameter estimates (1 PSE, 1 JND and 1 guess/lapse rate). Full model (3 PF model): three different PFs that differ only in their PSE values resulting in five parameter estimates (3 PSEs, 1 JND and 1 guess/lapse rate). Relative AIC: Akaike Information Criterion at the group level (subject level AICs summed up over subjects) relative to the null model (AIC = LL – 0.5 M; LL = log-likelihood, M = number of parameters). Greater relative AIC indicates a better fit of the model to the data. Exp. post. prob. = expected posterior probability; Exceedance prob. = exceedance probability; Prot. exceedance prob. = protected exceedance probability.

In addition, as a measure of participants' performance we calculated the hit rate and d-prime (d') for detecting the dimming fixation cross (in catch trials during pretest and posttest) or the visual stimuli (in catch trials during adaptation). In general, participants' performance detecting contrast changes was high throughout all three experiments (>85%; Table 6.2).

		Hit rate	d-prime	PSE	
Pretest					
	Behav	93.1% (±1.1%)	4.08 (±0.19)	0.59 (±0.22)	
	fMRI	97.5% (±1.0%)	4.80 (±0.29)	-0.72 (±0.48)	
	EEG	91.3% (±2.7%)	4.09 (±0.20)	-0.34 (±0.21)	
Adaptation					
AV-adaptation	Behav	96.4% (±1.4%)	4.65 (±0.21)		
	fMRI	92.4% (±1.1%)	4.54 (±0.08)	not applicable	
	EEG	89.6% (±4.0%)	4.38 (±0.53)		
VA-adaptation	Behav	95.6% (±1.6%)	4.69 (±0.22)		
	fMRI	89.5% (±2.7%)	4.09 (±0.23)		
	EEG	88.1% (±2.8%)	4.21 (±0.26)		
Posttest					
AV-posttest	Behav	91.4% (±1.7%)	4.07 (±0.20)	-0.95 (±0.29)	
	fMRI	97.8% (±0.6%)	5.01 (±0.23)	-3.34 (±0.21)	
	EEG	90.8% (±3.4%)	4.53 (±0.21)	-2.33 (±0.33)	
VA-posttest	Behav	91.2% (±1.4%)	4.09 (±0.19)	2.10 (±0.25)	
	fMRI	98.1% (±0.6%)	5.18 (±0.30)	1.66 (±0.54)	
	EEG	88.8% (±4.3%)	4.31 (±0.37)	1.87 (±0.41)	

Table 6.2 Behavioural results of hit rate, d' and PSE in the sound localization and adaptation tasks (group mean \pm SEM).

Multivariate decoding results

We used a multivariate decoding approach to characterize how the brain represents auditory spatial information across space and time. To map auditory space onto the patterns of the fMRI BOLD response and the EEG evoked potentials respectively, we trained a linear support vector regression model (SVR, LIBSVM (Chang & Lin, 2011)) separately for five pre-defined regions of interest (ROI) along the auditory spatial processing hierarchy (fMRI) and each time point up to 500 ms post stimulus (EEG). SVR models were trained to learn the mapping from activation patterns to the spatial location of the auditory stimuli using examples from the pretest conditions, then the learnt mapping was used to decode the spatial location from the BOLD response (fMRI) and evoked potential (EEG) patterns from the examples of the pretest and posttest phases.



Figure 6.5 Multivariate decoding results. (A) Bar graphs show group mean decoding accuracy for space coding (left bars) and recalibration index (right bars) across the ROIs from fMRI. Error bars represent SEM, (*: $p \leq 0.05$; **: $p \leq 0.01$). (B) Temporal dynamics of group mean decoding accuracy for space coding (top) and recalibration index (bottom) from EEG. The shaded area around the curves represents SEM, the shaded rectangles represent time windows with a significant effect. Abbreviations: HG: Heschl's gyrus; hA: higher auditory cortex; IPL: inferior parietal lobule; IPS: intraparietal sulcus; FEF: frontal eye-field; RI: recalibration index.

To decide whether the location of the signal could be successfully decoded from the fMRI and EEG activation patterns we assessed the accuracy to which the models predicted the true signal locations separately for each ROI (fMRI) and each time
point (EEG). For fMRI the following ROIs were investigated: Heschl's gyrus (HG), higher auditory cortex (hA, including planum temporale), inferior parietal lobule (IPL), intraparietal sulcus (IPS) and frontal eye-field (FEF). Decoding accuracy was evaluated as the Fisher z-transformed Pearson correlation coefficient between the true and decoded auditory locations for the pretest examples. The Fisher ztransformed correlation coefficients for every ROI (fMRI) and time point (EEG) were subsequently fed into a second level bootstrap-based one sample t-test. A cluster size based correction method was used to control the false positive rate for multiple comparisons (Nichols & Holmes, 2003) in the time domain. For fMRI, in line with previous research, the decoding accuracy for auditory locations was highest in planum temporale (Callan et al., 2015; Ortiz-Rios et al., 2017), but all ROIs displayed significantly better than chance performance (HG: p = 0.028; hA: p = 0.012; IPL: p =0.010; IPS: p = 0.005; FEF: p = 0.019) suggesting a network of brain regions sensitive to auditory spatial information (Figure 6.5A). The auditory spatial information could also be significantly decoded in the time domain from the EEG recordings. The decoding accuracy fluctuated around chance level (R = 0) until 100 ms post stimulus, when it started to rise steeply. Then the curve became less steep at around 145 ms post stimulus and reached its maximum at 355 ms (R = 0.39). Cluster based bootstrap test revealed one significant cluster (p = 0.010) starting at 110 ms post stimulus up until the end of the time window of interest, 500 ms (Figure 6.5B).

To investigate whether the representation of auditory space in the brain changes after exposure to spatially discrepant audiovisual stimulation we used the decoded auditory locations from the posttest examples. Specifically, we computed the

recalibration index (RI) for each ROI (fMRI) and time point (EEG) (see Methods). If there is a shift in the decoded auditory locations towards the visual stimulus, a RI value significantly greater than zero is expected. We found significant RI in HG (p =0.025), hA (p = 0.002), IPL (p = 0.01) and FEF (p = 0.037) (Figure 6.5A). The time course of the RI showed a sustained positive effect starting around 75 ms post stimulus. For statistical testing we restricted the time window of interest to the time window where we observed significant space coding (110 – 500 ms post stimulus). We found significant RI over the time window of interest as indicated by two significant clusters between 185 – 285 ms (p = 0.019) and 335 – 470 ms post stimulus (p = 0.005) (Figure 6.5B). These results collectively suggest that there is a system of brain regions involved both in the coding and in the recalibration of auditory space.

Characterization of neural representations of space and recalibration

As a first step to characterize the neural representations underlying the coding and recalibration of auditory space we used neurometric functions (NF). Neurometric functions are essentially psychometric functions, but relying on data from neural rather than behavioural responses. We hypothesised that brain regions sensitive to auditory space coding and recalibration would show neurometric functions with significant goodness-of-fit as a sign of neural representation underlying the behavioural responses.

We fitted neurometric functions to the percentage 'decoded right responses' as a function of stimulus location using binarized decoded spatial locations of pre- and posttest fMRI responses. We note that due to the lower signal-to-noise ratio of fMRI data, NFs were fitted on data averaged across all participants. In accordance with

our hypothesis, all tested ROIs showed neurometric functions with statistically significant goodness-of-fit (Figure 6.6A). Similarly to the behavioural results, recalibration shifted the NFs rightward after VA-adaptation (due to fewer decoded right responses) and leftward after AV-adaptation (due to more decoded right responses) with respect to the pretest NFs. Model comparison with likelihood-ratio test revealed that the full model (3 PF model) allowing for recalibration effects explained the data better than the null model (1 PF model) in all tested ROIs (HG: p = 0.0001; hA: p = 0.0001; IPL: p = 0.0001; IPS: p = 0.0001; FEF: p = 0.005; Figure 6.6A).



Figure 6.6 Characterization of neural representations – fMRI. (A) Neurometric function fits to the decoded auditory location in pretest, VA- and AV-posttest conditions across ROIs. (B) Space and recalibration representational dissimilarity matrices across ROIs and their projections into one dimension using multidimensional scaling (bottom). Abbreviations: HG: Heschl's gyrus; hA: higher auditory cortex; IPL: inferior parietal lobule; IPS: intraparietal sulcus; FEF: frontal eye-field;

To further characterize neural representations, we used representational similarity analysis (RSA) (Kriegeskorte, 2008) as implemented in the rsatoolbox (Nili et al., 2014). Representational dissimilarity matrices (RDM) were computed separately for each ROI (fMRI) and time point (EEG) using the Mahalanobis distance as the dissimilarity measure. For space coding, the RDMs were computed across the 7 spatial locations using data from the pretest trials, yielding 7x7 RDMs (Figure 6.2B left). For recalibration coding RDMs were computed across the 7-7 VA and AV posttest conditions, generating 14x14 RDMs (Figure 6.2B right). RDMs predicted by the population rate code model (Salminen et al., 2009; Trapeau & Schönwiesner, 2015) were also computed both for space and recalibration coding (Figure 6.2C).

The RDMs convey the similarity information in a complex high dimensional structure, which might be overwhelming to grasp at the first sight. To illustrate the similarity structure more intuitively we used multidimensional scaling (MDS) to project the data into a lower dimensional space (Figure 6.2C and 6.6).

First, the representational structures predicted by the population rate code model were examined. For space coding, dissimilarity increased monotonically with spatial disparity, i.e. it was greatest between the two extreme locations (-12° vs 12°), which

was also illustrated by MDS (Figure 6.2C bottom left). For recalibration coding the representational structure was identical to the space coding within both posttest conditions, however between posttest conditions the minimal dissimilarity shifted from the diagonal towards the centre of the RDM (Figure 6.2C top right). This is consistent with a shift of the representation towards the visual signal as illustrated by MDS as well (Figure 6.2C bottom right).

Then we tested whether the representation structures of space and recalibration coding observed in fMRI and EEG (Figure 6.6B) show similarity to the predictions of the population rate code model (Ortiz-Rios et al., 2017; Salminen, Tiitinen, Yrttiaho, & May, 2010; Trapeau & Schönwiesner, 2015). For this we computed the Spearman correlation coefficient between the model RDMs and the RDMs in each ROI and time point, both for space and recalibration coding. In accordance with the previous literature all ROIs showed significant relatedness to the model RDMs for space coding (HG: mean = 0.42, SEM = 0.29, p = 0.005; hA: mean = 1.21, SEM = 0.21, p = 0.002; IPL: mean = 0.94, SEM = 0.49, p = 0.011; IPS: mean = 1.11, SEM = 0.66, p = 0.010; FEF: mean = 0.69, SEM = 0.40, p = 0.010). We also found significant relatedness in the EEG time course as demonstrated by a significant cluster (p < r0.001) spanning from 85 ms to 500 ms p (Figure 6.7A bottom). Similarly, for recalibration coding, all ROIs showed significant relatedness to the model (HG: mean = 0.24, SEM = 0.13, p = 0.005; hA: mean = 0.68, SEM = 0.12, p = 0.010; IPL: mean = 0.40, SEM = 0.20, p = 0.010; IPS: mean = 0.53, SEM = 0.28, p = 0.010; FEF: mean = 0.30, SEM = 0.18, p = 0.027) as well as the EEG time course (significant cluster between 85 ms and 500 ms, p < 0.001) (Figure 6.7B bottom).

Furthermore, we asked how representations observed with fMRI corresponded to those observed with EEG. To investigate this we computed Spearman's correlation coefficient separately for each ROI between the RDM of the respective ROI and the RDM from each time point in the EEG time course. For space coding we found strong, sustained positive correlation between fMRI and EEG from ~85 ms until 500 ms post stimulus in hA, IPL, IPS and FEF, which has been confirmed by the bootstrap based statistical test (p < 0.001 for all) (see Figure 7A top). In HG the correlation was less strong, yet still significant, but notably more transient (significant cluster from 100 ms to 305 ms post stimulus, p = 0.004). We observed significant positive correlation between RDMs from fMRI and EEG in the recalibration coding in all ROIs except FEF. However, the time course of similarity showed a more sequential pattern across ROIs (Figure 6.7B top). HG showed the earliest cluster between 55 - 145 ms post stimulus (p = 0.04), followed by a more sustained cluster in hA between 130 – 435 ms (p < 0.001). Two significant clusters emerged in IPL in close proximity between 135 - 220 ms (p = 0.04) and between 245 - 355 ms (p = 0.027). Finally, one significant cluster emerged in IPS between 240 - 380 ms (p = 0.031). It should be noted that in FEF a marginally significant cluster emerged between 260 - 320 ms (p = 0.054).





Finally, we also performed a second order MDS analysis to illustrate the similarity between the representational structures in fMRI regions and models (Figure 6.8). For space coding HG proved to be separated from the rest of the regions which formed a

tight cluster with the model. hA and IPS showed the highest similarity with the model, followed by IPL and FEF. For recalibration coding hA and IPS once again showed the greatest similarity with the model, followed by IPL. FEF and HG proved to be separated from each other and the rest of the regions as well.



Figure 6.8 Second order MDS. Second order MDS plots showing the similarity between the ROIs and the models separately for space and recalibration coding.

Discussion

The nervous system must adapt to persistent changes that affect the sensory system at multiple timescales to maintain the consistency between sensory modalities and the accuracy of representations with respect to the physical properties of the world (Bur & Gori, 2012; De Gelder & Bertelson, 2003; Ernst & Di Luca, 2012). A vast body of behavioural research investigated spatial auditory adaptation induced by exposure to spatial audiovisual conflict, the so called ventriloquist aftereffect (Bertelson, Frissen, Vroomen, & De Gelder, 2006; Canon, 1970, 1971; L. Chen & Vroomen, 2013; Frissen et al., 2004, 2003; Lewald, 2002; M. Radeau & Bertelson, 1974; Monique Radeau & Bertelson, 1977; G. H. Recanzone, 1998; Wozny & Shams, 2011b). Despite the overwhelming amount of behavioural evidence, the underlying neural substrates remain elusive. The current study was designed to investigate the spatial (fMRI) and temporal (EEG) characteristics of the adaptive changes in the neural representations of auditory spatial processing induced by spatially discrepant visual signals. We used a classical recalibration paradigm with bisensory adaptation periods, preceded and followed by test phases of unisensory sound localization. Critically, participants performed adaptation periods with spatial discrepancy towards both directions (V stimulus to the left or right with respect to the A stimulus), which enabled us to compare recalibration effects both on the behavioural and neural levels without temporal confound. To minimize the influence of changes in motor responses on changes in the neural representation of auditory space, participants responded only on a fraction of trials in the sound localization task. These responses were used to confirm behavioural effects, whereas the rest of the trials were used to characterize the neural mechanisms of auditory spatial adaptation free from motor responses. It must be noted however, that motor preparation effects cannot be ruled out as participants localized the sounds on every trial and were prepared to respond regardless whether the response was required or not. On the behavioural level we found consistent recalibration effects in all three experiments (behavioural, fMRI, EEG). In particular, the point of subjective equality of the psychometric functions after adaptation moved to the direction as expected from the direction of the adaptation

consistently for every single participant (Figure 6.4B). Furthermore, Bayesian model comparison confirmed that a model with distinct PSE values for the conditions (pretest, AV- and VA-posstest) explained the data better than a model with a single, joint PSE value.

At the neural level, we first established whether auditory space can be decoded from the predefined ROIs along the auditory processing hierarchy (fMRI) and from the time course up to 500 ms post stimulus (EEG). The candidate regions of interest were: Heschl's gyrus (HG), higher auditory cortex (hA, including planum temporale), inferior parietal lobule (IPL), intraparietal sulcus (IPS) and frontal eye-field (FEF). These regions of interests were suggested both by behavioural studies investigating recalibration effects (Bruns & Röder, 2015; Frissen et al., 2004; Lewald, 2002; G. H. Recanzone, 1998) as well as neuroimaging studies characterizing auditory spatial processing in humans (Arnott et al., 2004; Bushara et al., 1999; Galperin, Ratner, & Nitzan, 2004; Michalka et al., 2016; Weeks et al., 1999). We trained linear support vector regression models for each subject on the auditory localization trials prior to adaptation to establish a mapping from BOLD- and EEG-response patterns to spatial locations of the auditory space. We then examined how well this mapping predicted the actual spatial locations. In line with previous research implicating that planum temporale has a prominent role in representing auditory space (Callan et al., 2015; Derey, Valente, De Gelder, & Formisano, 2016; Krumbholz et al., 2005; Porter, Semenov, Hulme, Harrison, & Arnell, 1999), this region showed the highest decoding accuracy. Interestingly, all other regions showed significantly better than chance decoding performance, even the primary auditory cortex (HG). This suggests that a network of regions is involved in the representation of auditory space. This is

supported by a clear evidence from a very recent fMRI study in macaque showing that hemifield code of the auditory space extends to the primary auditory cortices (Ortiz-Rios et al., 2017). Correspondingly, we found that auditory spatial representations were maintained for a sustained period starting from 110 ms until the end of the examined time window. This further corroborates that brain regions across multiple hierarchical levels represent auditory space. It is possible, that the results are affected by stimulus or working memory related effects. Further studies need to investigate to what extent the representation of auditory space is preserved in these regions under tasks that do not rely on spatial localization (e.g. passive listening).

Then we asked if signs of auditory space adaptation could be traced at the neural level. For this we used the previously learnt mappings from BOLD- and EEGresponse patterns to the auditory signals, and applied them to response patterns after visually induced spatial adaptation. Comparing the decoded spatial locations in AV- and VA-posttest conditions we assessed the recalibration index which shows the extent to which spatial recalibration occurred in the given ROI or time point. We found the strongest effect in PT, but to a lesser extent other regions also showed significant effects. This has been further corroborated by neurometric functions (cumulative Gaussians) fitted to the decoded locations (Figure 6.6A). EEG response patterns also reflected auditory space adaptation after 185 ms as indicated by two significant windows in close proximity. The first window between 185 - 285 ms coincides with the formation of the forced fusion spatial estimate of audiovisual signals (see Figure 5.7) pointing towards a potential link between mechanisms of integration (ventriloquist audiovisual effect) and audioviusal recalibration (ventriloquist aftereffect). We speculate that this potential link could be IPS as it has

been shown that the forced fusion estimate is predominantly represented by posterior IPS (Rohe & Noppeney, 2015a). Collectively, our findings suggest that recalibration induces changes in neural representations of auditory space across multiple levels of the hierarchy including unisensory auditory regions as hA. Interestingly, Bruns and colleagues (Bruns, Liebnau, & Röder, 2011) found ERP correlates of auditory space recalibration as early as 100 ms after stimulus onset: a significant interaction emerged between stimulus location and adaptation direction (i.e. leftward adaptation reduced the N100 component for right sounds, conversely rightward adaptation reduced the N100 component for left sounds). In the recalibration index dynamics we also observed a peak around this time; however the effect did not survive multiple comparisons. This might be because the RI takes all signal locations into consideration at once; consequently it is less sensitive to such minor changes. On the other hand, the aforementioned study failed to detect any later recalibration effect. This difference could be a consequence of their less ideal experimental design (i.e. adaptation direction as a between-subjects factor) or the increased sensitivity of our multivariate decoding approach.

Furthermore, we characterized the representations underlying auditory space and recalibration coding using representational similarity analysis (Kriegeskorte, 2008). First we compared the observed representations to the predictions of the population rate code model (Salminen et al., 2009). The neural representations in all ROIs and all time points after 85 ms were significantly related to the population rate code model predictions both for space and recalibration coding (Figure 6.6B), further corroborating previous evidence that auditory space is coded using a population rate code in the neocortex (Ortiz-Rios et al., 2017; Salminen et al., 2009; Trapeau &

Schönwiesner, 2015). Second, we compared the representations observed with the two imaging modalities, to obtain a unified picture of the spatio-temporal characteristics of auditory space and recalibration coding. The representations of space coding emerged in all regions from around 90 ms after stimulus onset and were maintained until the end of the time window, except for primary auditory cortex, where representations were more transient (Figure 6.7A). For recalibration coding we observed a more sequential evolution of representations (Figure 6.7B): first HG showed transient signs of recalibration coding (between 55 - 145 ms, however it must be noted that this time window largely precedes the time when auditory space or recalibration becomes decodable). This was followed by sustained representations in hA and in IPL, and finally in IPS and possibly in FEF (however correlations here were only marginally significant). The distinct spatio-temporal evolution of space and recalibration coding might indicate distinct underlying neural mechanisms: for space coding, there seems to be a uniform representation across ROIs (except for HG), whereas for recalibration coding representations are more specific to ROIs and emerge sequentially. A second order MDS analysis visualizing the similarity between the representations across ROIs further corroborated these speculations: for space coding, higher order ROIs were clustered closely, far apart from HG. For recalibration coding, ROIs dispersed indicating specific individual were more more representations.

Collectively our results demonstrate that auditory space and its adaptational changes are encoded in a hierarchy of cortical regions in the human neocortex including higher auditory cortices.

CHAPTER 7: GENERAL DISCUSSION

The work presented in this thesis aimed to further our understanding of the processes of multisensory integration and adaptation. In particular we asked if multisensory interactions occur prior to perceptual awareness (Chapter 4). We examined the temporal dynamics of the evolution of the estimates of Bayesian Causal Inference (Chapter 5) and the spatio-temporal characteristics of auditory space coding and recalibration (Chapter 6) in the human brain. In this chapter I summarize the main findings of each empirical chapter, show how they contribute to our current understanding of multisensory processing and outline directions for future research.

Summary of main findings

Chapter 4: A spatially collocated sound thrusts a flash into awareness

To interact effectively with the environment the brain integrates signals from multiple senses. It is currently unclear to what extent spatial information can be integrated across different senses in the absence of awareness. To investigate whether and how signals from different sensory modalities can interact prior to perceptual awareness, this chapter combined spatial audiovisual stimulation and continuous flash suppression.

We found that participants were more likely to detect the suppressed flash, when the flash was co-localized than non-collocated with the sound. This result suggest that an aware auditory signal can boost a weak visual signal into awareness in support of an 'automatic' account of audiovisual integration (Onat et al., 2007). Since audiovisual stimuli were kept in exact temporal synchrony, they also suggest that audiovisual

interactions prior to perceptual awareness are governed not only by temporal (as shown by Alsius and Munhall, 2013) but also by spatial constraints. Furthermore, these spatial constrains are likely to influence integration already in primary visual cortices as continuous flash suppression is thought to affect visual perception already at this early stage by attenuating neural activity similar to reducing the contrast of the stimulus (Yuval-Greenberg & Heeger, 2013).

Participants also located the flash in addition to judging its visibility on each trial, enabling us to evaluate the influence of the sound location on participants' perceived flash location depending on flash visibility. We expected that a spatially discrepant sound biases the perceived visual location as the spatial discrepancy was approximately 8 degrees visual angle, which is within the spatial window of integration. We observed that the influence of sound location on perceived flash location increased gradually from visible to unsure and invisible trials. This audiovisual spatial bias profile is consistent with the principle of reliability-weighted integration where a stronger weight should be given to the more reliable signal (Alais & Burr, 2004). It must be noted however, that our paradigm did not manipulate signal reliability, in contrast, the flashes were physically identical across all visibility levels. Yet, identical physical signals will elicit neural representations that vary in their reliability across trials because of trial-specific internal systems noise (Faisal et al., 2008). As the brain does not have access to the true physical reliability of the sensory signals but only to the uncertainty of the internal representation, our design still manipulated the reliability of the internal representations, Thus, our findings suggest that the relative auditory weight in the integration process depends on the reliability of the trial-specific internal representation evoked by the visual signal.

Chapter 5: To integrate or not to integrate: Temporal dynamics of Bayesian Causal Inference

The brain solves two fundamental computational problems when encountering conflicting multisensory signals: First, it determines whether sensory inputs originate from common or independent sources, the so-called causal inference problem. Second, the brain uses all available information to form the most reliable integrated percept if they are caused by a common source and, segregates them if signals originate from independent sources. Hierarchical Bayesian Causal Inference provides a unifying framework to incorporate both sensory integration and segregation in perception and indeed behavioural (Körding et al., 2007; Rohe & Noppeney, 2015b; Wozny et al., 2010) and neuroimaging (Rohe & Noppeney, 2015a, 2016) studies have demonstrated, that humans integrate audiovisual signals according to the predictions of this model. Importantly, the neuroimaging studies by Rohe and Noppeney used fMRI to investigate the processes of Bayesian Causal Inference across a hierarchy of brain regions. Yet, given the limited temporal resolution of fMRI, the temporal dynamics of multisensory perception and Bayesian Causal Inference are still unknown. To address these questions we adopted the experimental paradigm of Rohe and Noppeney (Rohe & Noppeney, 2015a, 2016) and combined it with EEG, multivariate pattern decoding and models of Bayesian Causal Inference.

Our results demonstrate that multisensory representations corresponding to the hierarchy of the Bayesian Causal Inference model emerge sequentially: First, until about 100 ms after stimulus onset, audiovisual signals are represented assuming

independent causes ('full segregation'). Similarly, the neural relative audiovisual weight w_{AV} showed visual dominance until 90 ms with limited auditory influence depending on spatial disparity providing corroborative evidence for inter-sensory interactions at the primary sensory cortical level (Kawecki & Ebert, 2004).

Then, between 100 - 290 ms, audiovisual stimuli are represented as if they were caused by a common cause ('forced fusion') weighted by their relative reliabilities. Indeed, neural w_{AV} showed influence of visual reliability on the representations from 90 ms. In line with this finding, it has been previously shown, the ventriloquist illusion is associated with the N260 component (Bonath et al., 2007), which is thought to be the result of reliability weighted integration (Alais & Burr, 2004).

Finally, the Bayesian Causal Inference estimate emerged after 290 ms. Similarly, the w_{AV} dynamics showed the incorporation of both task-relevance and relative sensory reliability into neural representation from 220 – 450 ms. Participants behavioural weights emerged in the brain between 340 – 495 ms coinciding with the Bayesian Causal Inference estimate which best explained the behaviour.

Collectively these results expand our understanding of how multisensory computations emerge across a hierarchy of brain regions.

Chapter 6: Spatio-temporal characteristics of visually induced auditory space adaptation in the human brain

Chapter 6 combined psychophysics, functional magnetic resonance imaging (fMRI), electroencephalography (EEG) and advanced multivariate decoding models to investigate the spatial (fMRI) and temporal (EEG) characteristics of visually induced

auditory adaptation in the human neocortex. Neural responses were examined in five predefined ROIs along the auditory processing hierarchy (fMRI) and each time point up to 500 ms post stimulus (EEG). The candidate regions of interest were: Heschl's gyrus (HG), higher auditory cortex (hA, including planum temporale), inferior parietal lobule (IPL), intraparietal sulcus (IPS) and frontal eye-field (FEF).

First, we established that auditory space can be decoded from neural activation patterns. As anticipated based on previous evidence (Callan et al., 2015; Derey et al., 2016; Krumbholz et al., 2005; Porter et al., 1999), planum temporale showed the highest decoding accuracy, but all other regions showed significantly better than chance decoding performance, suggesting the involvement of a network of regions in the representation of auditory space. Correspondingly, auditory spatial representations were maintained throughout the trial starting from 110 ms post stimulus. This extends previous evidence from nonhuman primates (Ortiz-Rios et al., 2017) to humans, that hemifield code of the auditory space extends to the primary and higher auditory cortices.

Then we asked if auditory space adaptation could be decoded from the neural responses. We found the strongest effect in hA, but to a lesser extent other regions also showed significant effects, except for IPS, where a trend towards significance was observed. EEG response patterns also reflected auditory space adaptation after 185 ms. Our findings suggest that spatially discrepant visual signals induce changes in neural representations of auditory space across multiple levels of the hierarchy including unisensory auditory regions as hA probably through IPS.

We used representational similarity analysis to characterize the neural responses underlying auditory space and recalibration coding. First we assessed the relatedness of observed representations to the predictions of the population rate code model (Salminen et al., 2009). All ROIs and all time points after 85 ms exhibited representations consistent with the population rate code model, further corroborating previous evidence that auditory space is coded according to a population rate code in the neocortex (Ortiz-Rios et al., 2017; Salminen et al., 2009; Trapeau & Schönwiesner, 2015). Second, we compared the representations across the two imaging modalities obtaining a unified, spatio-temporal view on auditory space and recalibration coding. For space coding a uniform representation emerged across ROIs (except for HG), from around 90 ms after stimulus onset and was maintained until the end of the time window. For recalibration coding representations proved to be more specific to ROIs and emerged sequentially from lower level to higher level areas. The distinct spatio-temporal evolution of space and recalibration coding might indicate distinct underlying neural mechanisms, which was further corroborated by a second order MDS analysis.

Contributions and future implications

In *Chapter 4* we demonstrated that participants were more likely to detect a suppressed flash, when it was co-localized than non-collocated with the sound. As we discussed earlier, this can be the result of both top-down and bottom-up mechanisms such as spatial co-occurrence. Indeed, spatial congruency is thought to act as a fundamental principle of multisensory integration (i.e. the spatial rule (Stein & Meredith, 1993)), however its importance has been questioned recently (Spence,

2013). Our results in Chapter 5 demonstrate that spatial disparity influences the integration of audiovisual signals early, up to 100 ms after stimulus onset, pointing towards the involvement of primary sensory areas. This was also explicitly demonstrated with fMRI by Rohe and Noppeney (Rohe & Noppeney, 2016) supporting the role of bottom-up mechanisms behind the boosting of suppressed visual signals into perceptual awareness depending on spatial congruency. However, future electrophysiological studies could much more precisely delineate between topdown and bottom-up mechanisms. Also, future studies could further characterize the critical spatial window of integration under flash suppression by systematically manipulating the spatial discrepancy of the audiovisual signals. Combined with EEG or fMRI this research line would allow the precise description of the cortical levels where sounds interact with visual processing under flash suppression. Furthermore our results in Chapter 4 showed that the influence of sound location on perceived flash location increased as the visibility of the trials decreased, which we hypothesised to be the result of dynamic reweighting of sensory signals due to decreasing reliability of internal visual representations (Alais & Burr, 2004). As this study did not manipulate external reliability of sensory stimuli it would be interesting to conduct a neuroimaging experiment using this paradigm and compare the results to Chapter 5 where external stimulus reliability was explicitly manipulated.

In *Chapter 5* we demonstrated that multisensory interactions evolve dynamically: different computational operations emerge sequentially in line with the hierarchy of the Bayesian Causal Inference model. To our knowledge this is the first study to examine the temporal dynamics of Bayesian Causal Inference in the brain. A step forward could be to temporally resolve these processes in separate brain regions.

One way to achieve this could be to decode from source-reconstructed data. Unfortunately our methodology was limited in this respect (64-channel EEG) and we could not reliably decode in source space. In the future a similar experiment could be performed with MEG which has a much finer spatial resolution because of more detectors (>300, depending on setup) and the signals are not smeared by the skull and the scalp. In our experimental design the effect of task relevance was kept constant within each run, hence participants could anticipate the modality of the stimulus. It would be interesting to see whether the time course of the task relevance effect would change if participants could not anticipate the to-be-responded modality. A similar experiment with randomized task relevance condition could answer this question. Finally it would be interesting to examine the neural mechanisms of BCI in other experimental paradigms. One such candidate could be the sound induced flash illusion, where a single flash accompanied by multiple auditory beeps is perceived as multiple flashes (Shams, Ma, & Beierholm, 2005). Importantly, in the temporal modality hearing is more reliable than vision; hence in this task hearing dominates the percept in contrast to the ventriloguist task. Future neuroimaging experiment could investigate the neural mechanisms of BCI in the context of the sound induced flash illusion as well.

In *Chapter 6* we shed light on the neural processes behind multisensory recalibration and showed that there is a network of brain regions involved in the representation and adaptation of auditory space which includes even the primary auditory cortex. In a very recent fMRI experiment Zierul and colleagues (2017) reached very similar conclusions: their results suggest that adaptation occurs in low-level auditory areas mediated by spatial representations in parietal cortex. Importantly, we demonstrated

that hA is involved in the representation of auditory space, furthermore these representations are sensitive to spatially discrepant visual stimuli and recalibrated, probably mediated through IPS. An outstanding question is whether the spatial representations found in these cortical areas are dependent on task or memory effects. Further studies are needed to investigate this question where the spatial auditory stimulation does not involve localization of the stimuli (e.g. passive listening).

In our experiment, we separated the adaptation and the test periods and interleaved them to make sure we elicit an effect which is as robust as possible. Recent studies demonstrated, that adaptive processes can take place much more rapidly (Frissen et al., 2012; Mendonca et al., 2015; Wozny & Shams, 2011a), even after exposure to a single spatially discrepant audiovisual stimulus pair (Wozny & Shams, 2011b). Future studies with interleaved adaptation trials could investigate how these rapid, trial-bytrial adaptation processes evolve across time and space in the human brain.

According to recent theories of multisensory recalibration both supervised (i.e. with external feedback) and unsupervised (without external feedback) calibration act together simultaneously, with the former restoring the accuracy of the multisensory percept and the latter minimizing the discrepancy between the sensory modalities (A. Zaidel et al., 2011; Adam Zaidel et al., 2013). In our experiment in *Chapter 5* we did not give external feedback to participants so our results reflect the neural substrates of unsupervised recalibration. It is an interesting question whether these processes (i.e. supervised vs. unsupervised) have different neural underpinnings. Importantly, it has been demonstrated that supervised recalibration depends on stimulus reliability. A future experiment using our paradigm with feedback and manipulating visual

reliability systematically may investigate the neural mechanisms of supervised recalibration and its dependence on stimulus reliability.

The ventriloquist aftereffect is a rapid visual-induced recalibration, however auditory space adaptation can be elicited by other means, for example applying ear moulds or earplugs, altering HRTFs or using electronic hearing devices (Mendonca et al., 2015). Our experimental approach could be utilized to compare the neural mechanisms of visually induced space adaptation to the ones elicited using the above methods. Furthermore, in a future experiment the time course of adaptation could also be altered from minutes to days such as in the study of Trapeau and Schönweisner (2015) and could lead to interesting comparisons between the spatiotemporal dynamics of rapid vs. long-term adaptation of auditory space.

Conclusions

Our experiments in the present thesis investigated multisensory integration and adaptation using a wide range of methods. The overarching theme across all of the studies was space: the spatial aspects of multisensory integration and how the brain represents and recalibrates auditory space. We demonstrated that degraded visual stimuli may or may not enter consciousness depending on spatial correspondences with a concurrent sound indicating that audiovisual integration is preserved inside the spatial window of integration even in the absence of awareness. We have shown the dynamics of how the brain takes into consideration the task relevance and reliability of signals when they are integrated and crucially the temporal aperture where this integration breaks down if the stimuli are outside the spatial window of integration. This window between roughly 300 – 500 ms exhibits neural representations predicted by the Bayesian Causal Inference model which weights integration and segregation

estimates according to their most likely causal structure. Finally we demonstrated how auditory space is represented across a network of brain regions and how consistent spatially discrepant audiovisual stimulation leads to recalibration in these areas. Furthermore, the time course of auditory space recalibration and the fusion of audiovisual signals roughly overlaps (200 – 300 ms) pointing towards similar mechanisms underlying spatial recalibration and integration.

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
- Adams, W. J., Banks, M. S., & Van Ee, R. (2001). Adaptation to three-dimensional distortions in human vision. *Nature Neuroscience*, *4*(11), 1063–1064. https://doi.org/10.1038/nn729
- Akaike, H. (1974). A New Look at the Statistical Model Identification. *IEEE Transactions on Automatic Control*, *19*(6), 716–723. https://doi.org/10.1109/TAC.1974.1100705
- Alais, D., & Burr, D. (2004). Ventriloquist Effect Results from Near-Optimal Bimodal Integration. *Current Biology*, 14(3), 257–262. https://doi.org/10.1016/S0960-9822(04)00043-0
- Alais, D., van Boxtel, J. J., Parker, A., & van Ee, R. (2010). Attending to auditory signals slows visual alternations in binocular rivalry. *Vision Research*, 50(10), 929–935. https://doi.org/10.1016/j.visres.2010.03.010
- Allefeld, C., Görgen, K., & Haynes, J. D. (2016). Valid population inference for information-based imaging: From the second-level t-test to prevalence inference. *NeuroImage*, 141, 378–392. https://doi.org/10.1016/j.neuroimage.2016.07.040
- 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), 16. https://doi.org/10.3389/fnint.2015.00016
- 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
- Alsius, A., Navarra, J., Campbell, R., & Soto-Faraco, S. (2005). Audiovisual integration of speech falters under high attention demands. *Current Biology*, *15*(9), 839–843. https://doi.org/10.1016/j.cub.2005.03.046
- Alsius, A., Navarra, J., & Soto-Faraco, S. (2007). Attention to touch weakens audiovisual speech integration. *Experimental Brain Research*, *183*(3), 399–404. https://doi.org/10.1007/s00221-007-1110-1
- Amedi, A. (2002). Convergence of Visual and Tactile Shape Processing in the Human Lateral Occipital Complex. *Cerebral Cortex*, 12(11), 1202–1212. https://doi.org/10.1093/cercor/12.11.1202
- Andersen, T. S., Tiippana, K., & Sams, M. (2004). Factors influencing audiovisual fission and fusion illusions. *Cognitive Brain Research*, 21(3), 301–308. https://doi.org/10.1016/j.cogbrainres.2004.06.004
- Anderson, C. M., & Wu, C. F. J. (1995). Measuring Location Effects from Factorial Experiments with a Directional Response. *International Statistical Review*, *63*(3),

345–363. Retrieved http://www.jstor.org/stable/1403484%5Cnhttp://about.jstor.org/terms

from

- Arndt, P. A., & Colonius, H. (2003). Two stages in crossmodal saccadic integration: Evidence from a visual-auditory focused attention task. *Experimental Brain Research*, 150(4), 417–426. https://doi.org/10.1007/s00221-003-1424-6
- Arnott, S. R., Binns, M. A., Grady, C. L., & Alain, C. (2004). Assessing the auditory dual-pathway model in humans. *NeuroImage*, 22(1), 401–408. https://doi.org/10.1016/j.neuroimage.2004.01.014
- Atkins, J. E., Jacobs, R. A., & Knill, D. C. (2003). Experience-dependent visual cue recalibration based on discrepancies between visual and haptic percepts. *Vision Research*, 43(25), 2603–2613. https://doi.org/10.1016/S0042-6989(03)00470-X
- Avillac, M., Ben Hamed, S., & Duhamel, J.-R. (2007). Multisensory Integration in the Ventral Intraparietal Area of the Macaque Monkey. *Journal of Neuroscience*, 27(8), 1922–1932. https://doi.org/10.1523/JNEUROSCI.2646-06.2007
- 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
- Barbas, H., Medalla, M., Alade, O., Suski, J., Zikopoulos, B., & Lera, P. (2005). Relationship of prefrontal connections to inhibitory systems in superior temporal areas in the rhesus monkey. *Cerebral Cortex*, 15(9), 1356–1370. https://doi.org/10.1093/cercor/bhi018
- Barraclough, N. E., Xiao, D., Baker, C. I., Oram, M. W., & Perrett, D. I. (2005). Integration of visual and auditory information by superior temporal sulcus neurons responsive to the sight of actions. *Journal of Cognitive Neuroscience*, 17(3), 377–391. https://doi.org/10.1162/0898929053279586
- Batteau, D. W. (1967). The Role of the Pinna in Human Localization. *Proceedings of the Royal Society B: Biological Sciences*, *168*(1011), 158–180. https://doi.org/10.1098/rspb.1967.0058
- Beauchamp, M. S. (2005). See me, hear me, touch me: Multisensory integration in lateral occipital-temporal cortex. *Current Opinion in Neurobiology*, 15(2), 145– 153. https://doi.org/10.1016/j.conb.2005.03.011
- Beauchamp, M. S., Argall, B. D., Bodurka, J., Duyn, J. H., & Martin, A. (2004). Unraveling multisensory integration: Patchy organization within human STS multisensory cortex. *Nature Neuroscience*, 7(11), 1190–1192. https://doi.org/10.1038/nn1333
- Beierholm, U. R., Quartz, S. R., & Shams, L. (2009). Bayesian priors are encoded independently from likelihoods in human multisensory perception. *Journal of Vision*, 9(5), 23–23. https://doi.org/10.1167/9.5.23
- Bertelson, P., & Aschersleben, G. (1998). Automatic visual bias of perceived auditory location. *Psychonomic Bulletin and Review*, *5*(3), 482–489. https://doi.org/10.3758/BF03208826

- Bertelson, P., Frissen, I., Vroomen, J., & De Gelder, B. (2006). The aftereffects of ventriloquism: Patterns of spatial generalization. *Perception and 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. (1981). Cross-modal bias and perceptual fusion with auditory-visual spatial discordance. *Perception & Psychophysics*, 29(6), 578– 584. https://doi.org/10.3758/BF03207374
- Bertini, C., Leo, F., & Làdavas, E. (2008). Temporo-nasal asymmetry in multisensory integration mediated by the Superior Colliculus. *Brain Research*, *1242*, 37–44. https://doi.org/10.1016/j.brainres.2008.03.087
- Bischoff, M., Walter, B., Blecker, C. R., Morgen, K., Vaitl, D., & Sammer, G. (2007). Utilizing the ventriloquism-effect to investigate audio-visual binding. *Neuropsychologia*, 45(3), 578–586. https://doi.org/10.1016/j.neuropsychologia.2006.03.008
- Bisley, J. W., & Goldberg, M. E. (2010). Attention, Intention, and Priority in the Parietal Lobe. *Annual Review of Neuroscience*, *33*(1), 1–21. https://doi.org/10.1146/annurev-neuro-060909-152823
- 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
- Bolognini, N., Frassinetti, F., Serino, A., & Làdavas, E. (2005). "Acoustical vision" of below threshold stimuli: Interaction among spatially converging audiovisual inputs. *Experimental Brain Research*, 160(3), 273–282. https://doi.org/10.1007/s00221-004-2005-z
- 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., & 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
- Boyle, S. C., Kayser, S. J., & Kayser, C. (2017). Neural correlates of multisensory reliability and perceptual weights emerge at early latencies during audio-visual integration. *European Journal of Neuroscience*, 46(10), 2565–2577. https://doi.org/10.1111/ejn.13724
- Brainard, D. H. (1997). The Psychophysics Toolbox. *Spatial Vision*, *10*(4), 433–436. https://doi.org/10.1163/156856897X00357
- Brand, A., Behrend, O., Marquardt, T., McAlpine, D., & Grothe, B. (2002). Precise inhibition is essential for microsecond interaural time difference coding. *Nature*,

417(6888), 543-547. https://doi.org/10.1038/417543a

- Bresciani, J.-P., Dammeier, F., & Ernst, M. O. (2006). Vision and touch are automatically integrated for the perception of sequences of events. *Journal of Vision*, *6*(5), 2. https://doi.org/10.1167/6.5.2
- Brosch, M. (2005). Nonauditory Events of a Behavioral Procedure Activate Auditory Cortex of Highly Trained Monkeys. *Journal of Neuroscience*, *25*(29), 6797–6806. https://doi.org/10.1523/JNEUROSCI.1571-05.2005
- Bruce, C., Desimone, R., & Gross, C. G. (1981). Visual properties of neurons in a polysensory area in superior temporal sulcus of the macaque. *Journal of Neurophysiology*, 46(2), 369–384. https://doi.org/10.1152/jn.1981.46.2.369
- Bruns, P., Liebnau, R., & Röder, B. (2011). Cross-Modal Training Induces Changes in Spatial Representations Early in the Auditory Processing Pathway. *Psychological Science*, 22(9), 1120–1126. https://doi.org/10.1177/0956797611416254
- Bruns, P., Maiworm, M., & Röder, B. (2014). Reward expectation influences audiovisual spatial integration. *Attention, Perception, and Psychophysics*, 76(6), 1815–1827. https://doi.org/10.3758/s13414-014-0699-y
- Bruns, P., & Röder, B. (2015). Sensory recalibration integrates information from the immediate and the cumulative past. *Scientific Reports*, *5*, 12739. https://doi.org/10.1038/srep12739
- Bur, D., & Gori, M. (2012). Chapter 18 Multisensory Integration Develops Late in Humans. In *The Neural Bases of Multisensory Processes* (pp. 1–18). https://doi.org/NBK92864 [bookaccession]
- Burenko, Y. A. (2010). Information component of managerial decision-making process concerning the development of processing enterprises. *Actual Problems of Economics*, *11*(9), 289–294. https://doi.org/10.1162/089892999563544
- Burge, J., Girshick, A. R., & Banks, M. S. (2010). Visual-Haptic Adaptation Is Determined by Relative Reliability. *Journal of Neuroscience*, *30*(22), 7714–7721. https://doi.org/10.1523/JNEUROSCI.6427-09.2010
- Bushara, K. O., Weeks, R. A., Ishii, K., Catalan, M. J., Tian, B., Rauschecker, J. P., & Hallett, M. (1999). Modality-specific frontal and parietal areas for auditory and visual spatial localization in humans. *Nature Neuroscience*, 2(8), 759–766. https://doi.org/10.1038/11239
- Busse, L., Roberts, K. C., Crist, R. E., Weissman, D. H., & Woldorff, M. G. (2005). The spread of attention across modalities and space in a multisensory object. *Proceedings of the National Academy of Sciences*, 102(51), 18751–18756. https://doi.org/10.1073/pnas.0507704102
- Butler, R. A. (1969). Monaural and binaural localization of noise bursts vertically in the median sagittal plane. *J. Audit. Res.*, *3*(3), 230–235.
- Butler, R. A. (1972). The influence of spatial separation of sound sources on the auditory evoked response. *Neuropsychologia*, 10(2), 219–225.

https://doi.org/10.1016/0028-3932(72)90063-2

- Calhoun, V. D., & Adali, T. (2006). Unmixing fMRI with independent component analysis. *IEEE Engineering in Medicine and Biology Magazine*, *25*(2), 79–90. https://doi.org/10.1109/MEMB.2006.1607672
- Callan, A., Callan, D., & Ando, H. (2015). An fMRI study of the ventriloquism effect. *Cerebral Cortex*, 25(11), 4248–4258. https://doi.org/10.1093/cercor/bhu306
- Calvert, G. A., Campbell, R., & Brammer, M. J. (2000). Evidence from functional magnetic resonance imaging of crossmodal binding in the human heteromodal cortex. *Current Biology*, *10*(11), 649–657. https://doi.org/10.1016/S0960-9822(00)00513-3
- Calvert, G. A., Hansen, P. C., Iversen, S. D., & Brammer, M. J. (2001). Detection of audio-visual integration sites in humans by application of electrophysiological criteria to the BOLD effect. *NeuroImage*, 14(2), 427–438. https://doi.org/10.1006/nimg.2001.0812
- Canon, L. K. (1970). Intermodality inconsistency of input and directed attention as determinants of the nature of adaptation. *Journal of Experimental Psychology*, *84*(1), 141–147. https://doi.org/10.1037/h0028925
- Canon, L. K. (1971). Directed attention and maladaptive "adaptation" to displacement of the visual field. *Journal of Experimental Psychology*, *88*(3), 403–408. https://doi.org/10.1037/h0030878
- 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., Rouiller, E. M., & Barone, P. (2012). Cortical and Thalamic Pathways for Multisensory and Sensorimotor Interplay. In *The Neural Bases of Multisensory Processes* (pp. 13–28). https://doi.org/doi:10.1201/b11092-4
- Carr, C. E., & Konishi, M. (1988). Axonal delay lines for time measurement in the owl's brainstem. *Proceedings of the National Academy of Sciences of the United States of America*, 85(21), 8311–8315. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/3186725
- Carr, C. E., & Konishi, M. (1990). A circuit for detection of interaural time differences in the brain stem of the barn owl. *The Journal of Neuroscience*, *10*(10), 3227– 3246. https://doi.org/10.1523/JNEUROSCI.10-10-03227.1990
- 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
- Chechik, G., & Nelken, I. (2012). Auditory abstraction from spectro-temporal features to coding auditory entities. *Proceedings of the National Academy of Sciences*, *109*(46), 18968–18973. https://doi.org/10.1073/pnas.1111242109
- Chen, L., & Vroomen, J. (2013). Intersensory binding across space and time: A tutorial review. *Attention, Perception, and Psychophysics*, 75(5), 790–811.

https://doi.org/10.3758/s13414-013-0475-4

- Chen, Y. C., & Spence, C. (2011). The Crossmodal Facilitation of Visual Object Representations By Sound: Evidence From the Backward Masking Paradigm. *Journal of Experimental Psychology: Human Perception and Performance*, 37(6), 1784–1802. https://doi.org/10.1037/a0025638
- Chen, Y. C., & Yeh, S. L. (2009). Catch the moment: Multisensory enhancement of rapid visual events by sound. *Experimental Brain Research*, *198*(2–3), 209–219. https://doi.org/10.1007/s00221-009-1831-4
- 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(SEP), 212. https://doi.org/10.3389/fpsyg.2011.00212
- Clemo, H., Keniston, L., & Meredith, M. A. (2011). Structural Basis of Multisensory Processing. In *The Neural Bases of Multisensory Processes* (pp. 3–14). https://doi.org/10.1201/b11092-3
- Conrad, V., Bartels, A., Kleiner, M., & Noppeney, U. (2010). Audiovisual interactions in binocular rivalry. *Journal of Vision*, *10*(10), 27–27. 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
- Dale, A. M., Fischl, B., & Sereno, M. I. (1999). Cortical Surface-Based Analysis: I. Segmentation and Surface Reconstruction. *NeuroImage*, 9(2), 179–194. https://doi.org/10.1006/nimg.1998.0395
- De Gelder, B., & Bertelson, P. (2003). Multisensory integration, perception and ecological validity. *Trends in Cognitive Sciences*, 7(10), 460–467. https://doi.org/10.1016/j.tics.2003.08.014
- De Leeuw, J. (2009). Journal of statistical software. *Wiley Interdisciplinary Reviews: Computational Statistics*, 1(1), 128–129. https://doi.org/10.1002/wics.10
- 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
- Derey, K., Valente, G., De Gelder, B., & Formisano, E. (2016). Opponent Coding of Sound Location (Azimuth) in Planum Temporale is Robust to Sound-Level Variations. *Cerebral Cortex*, 26(1), 450–464. https://doi.org/10.1093/cercor/bhv269
- 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(6), 585–606. https://doi.org/10.1163/22134808-00002529
- Destrieux, C., Fischl, B., Dale, A. M., & Halgren, E. (2010). Automatic parcellation of human cortical gyri and sulci using standard anatomical nomenclature. *NeuroImage*, *53*(1), 1–15. https://doi.org/10.1016/j.neuroimage.2010.06.010
- Diederich, A., Colonius, H., Bockhorst, D., & Tabeling, S. (2003). Visual-tactile spatial interaction in saccade generation. *Experimental Brain Research*, *148*(3), 328–337. https://doi.org/10.1007/s00221-002-1302-7
- Driver, J., & Noesselt, T. (2008). Multisensory Interplay Reveals Crossmodal Influences on "Sensory-Specific" Brain Regions, Neural Responses, and Judgments. *Neuron*, *57*(1), 11–23. https://doi.org/10.1016/j.neuron.2007.12.013
- Duhamel, J.-R., Colby, C. L., & Goldberg, M. E. (1998). Ventral Intraparietal Area of the Macaque: Congruent Visual and Somatic Response Properties. *Journal of Neurophysiology*, 79(1), 126–136. https://doi.org/10.1152/jn.1998.79.1.126
- Dumas, B., Lalanne, D., & Oviatt, S. (2009). Multimodal interfaces: A survey of principles, models and frameworks. *Lecture Notes in Computer Science* (*Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics*), 5440 LNCS(2), 3–26. https://doi.org/10.1007/978-3-642-00437-7_1
- Edgington, E. S., & Onghena, P. (2007). *Randomization Tests. SpringerReference* (4th ed.). Chapman & Hall/CRC. https://doi.org/10.1007/SpringerReference_205144
- Efron, B., & Tibshirani, R. J. (1993). *An Introduction to the bootstrap*. Boston, MA: Springer US. https://doi.org/10.1007/978-1-4899-4541-9
- 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
- Ernst, M. O., & Di Luca, M. (2012). Multisensory Perception: From Integration to Remapping. In Sensory Cue Integration (pp. 224–250). Oxford, UK: Oxford University Press. https://doi.org/10.1093/acprof:oso/9780195387247.003.0012
- Fairhall, S. L., & MacAluso, E. (2009). Spatial attention can modulate audiovisual integration at multiple cortical and subcortical sites. *European Journal of Neuroscience*, 29(6), 1247–1257. https://doi.org/10.1111/j.1460-9568.2009.06688.x

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., 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
- Felleman, D. J., & Van Essen, D. C. (1991). Distributed hierarchical processing in the primate cerebral cortex. *Cerebral Cortex*, 1(1), 1–47. https://doi.org/10.1093/cercor/1.1.1
- Fetsch, C. R., Deangelis, G. C., & Angelaki, D. E. (2013). Bridging the gap between theories of sensory cue integration and the physiology of multisensory neurons. *Nature Reviews Neuroscience*, 14(6), 429–442. https://doi.org/10.1038/nrn3503
- Fetsch, C. R., Pouget, A., Deangelis, G. C., & Angelaki, D. E. (2012). Neural correlates of reliability-based cue weighting during multisensory integration. *Nature Neuroscience*, 15(1), 146–154. https://doi.org/10.1038/nn.2983
- Fetsch, C. R., Turner, A. H., DeAngelis, G. C., & Angelaki, D. E. (2009). Dynamic Reweighting of Visual and Vestibular Cues during Self-Motion Perception. *Journal of Neuroscience*, 29(49), 15601–15612. https://doi.org/10.1523/JNEUROSCI.2574-09.2009
- Fisher, H. G., & Freedman, S. J. (1968). The role of the pinna in auditory localization. *Journal of Auditory Research*, 8(1), 15–26.
- Fishman, M. C., Fishman, M. C., Michael, P., & Michael, C. R. (1973). Integration of Auditory Information in the Cat's Visual Cortex. *Vision Research*, 13(8), 1415– 1419. https://doi.org/10.1016/0042-6989(73)90002-3
- Fogelson, S. V., Kohler, P. J., Miller, K. J., Granger, R., & Tse, P. U. (2014). Unconscious neural processing differs with method used to render stimuli invisible. *Frontiers in Psychology*, 5(JUN), 601. https://doi.org/10.3389/fpsyg.2014.00601
- Forster, B., Cavina-Pratesi, C., Aglioti, S. M., & Berlucchi, G. (2002). Redundant target effect and intersensory facilitation from visual-tactile interactions in simple reaction time. *Experimental Brain Research*, 143(4), 480–487. https://doi.org/10.1007/s00221-002-1017-9
- 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. https://doi.org/DOI 10.1152/jn.00694.2001
- Frassinetti, F., Bolognini, N., & Làdavas, E. (2002). Enhancement of visual perception by crossmodal visuo-auditory interaction. *Experimental Brain Research*, *147*(3), 332–343. https://doi.org/10.1007/s00221-002-1262-y
- Frissen, I., Vroomen, J., & De Gelder, B. (2012). The aftereffects of ventriloquism: The time course of the visual recalibration of auditory localization. *Seeing and Perceiving*, 25(1), 1–14. https://doi.org/10.1163/187847611X620883

- 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
- Frissen, I., Vroomen, J., de Gelder, B., Bertelson, P., Gelder, D., & Bertelson, P. (2004). The aftereffects of ventriloquism: generalization across soundfrequencies. Acta Psychologica, 118(1–2), 93–100. https://doi.org/10.1016/j.actpsy.2004.10.004
- Friston, K. J., Holmes, A. P., Poline, J.-B., Grasby, P. J., Williams, S. C., Frackowiak, R. S. J., & Turner, R. (1995). Analysis of fMRI time-series revisited. *NeuroImage*, 2(1), 45–53. https://doi.org/10.1006/nimg.1995.1007
- Friston, K. J., Holmes, A. P., Worsley, K. J., Frith, C. D., & Frackowiak, R. S. J. (1995). Statistical parametric maps in functional imaging: a general linear approach. *Human Brain Mapping*, 2, 189–210.
- Friston, K. J., Holmes, A. P., Worsley, K. J., Poline, J. -P, Frith, C. D., & Frackowiak, R. S. J. (1994). Statistical parametric maps in functional imaging: A general linear approach. *Human Brain Mapping*, 2(4), 189–210. https://doi.org/10.1002/hbm.460020402
- Friston, K. J., Jezzard, P., & Turner, R. (1994). Analysis of functional MRI timeseries. *Human Brain Mapping*, 1(2), 153–171. https://doi.org/10.1002/hbm.460010207
- Fu, K.-M. G. (2004). Timing and Laminar Profile of Eye-Position Effects on Auditory Responses in Primate Auditory Cortex. *Journal of Neurophysiology*, 92(6), 3522–3531. https://doi.org/10.1152/jn.01228.2003
- Galperin, M., Ratner, M. A., & Nitzan, A. (2004). Inelastic electron tunneling spectroscopy in molecular junctions: Peaks and dips. *Journal of Chemical Physics*, *121*(23), 11965–11979. https://doi.org/10.1093/cercor/bhs359
- 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
- Ghahramani, Z., Wolpert, D. M., & Michale, I. J. (1997). Computational models of sensorimotor integration. In *Advances in Psychology* (Vol. Volume 119, pp. 117–147). Retrieved from http://www.sciencedirect.com/science/article/pii/S0166411597800064
- Ghazanfar, A. A. (2005). Multisensory Integration of Dynamic Faces and Voices in Rhesus Monkey Auditory Cortex. *Journal of Neuroscience*, *25*(20), 5004–5012. https://doi.org/10.1523/JNEUROSCI.0799-05.2005
- Ghazanfar, A. A., & Schroeder, C. E. (2006). Is neocortex essentially multisensory? *Trends in Cognitive Sciences*, *10*(6), 278–285. https://doi.org/10.1016/j.tics.2006.04.008
- Girard, S., Collignon, O., & Lepore, F. (2011). Multisensory gain within and across hemispaces in simple and choice reaction time paradigms. *Experimental Brain*

Research, 214(1), 1–8. https://doi.org/10.1007/s00221-010-2515-9

- Gonzalez, L., & Manly, B. F. J. (1998). Analysis of variance by randomization with small data sets. *Environmetrics*, *9*(1), 53–65. https://doi.org/10.1002/(SICI)1099-095X(199801/02)9:1<53::AID-ENV285>3.0.CO;2-#
- Gori, M., Del Viva, M., Sandini, G., & Burr, D. C. (2008). Young Children Do Not Integrate Visual and Haptic Form Information. *Current Biology*, *18*(9), 694–698. https://doi.org/10.1016/j.cub.2008.04.036
- Gori, M., Sandini, G., Martinoli, C., & Burr, D. (2010). Poor Haptic Orientation Discrimination in Nonsighted Children May Reflect Disruption of Cross-Sensory Calibration. *Current Biology*, 20(3), 223–225. https://doi.org/10.1016/j.cub.2009.11.069
- Gottlieb, J. P., Kusunoki, M., & Goldberg, M. E. (1998). The representation of visual salience in monkey parietal cortex. *Nature*, *391*(6666), 481–484. https://doi.org/10.1038/35135
- Gramfort, A., Luessi, M., Larson, E., Engemann, D. A., Strohmeier, D., Brodbeck, C., ... Hämäläinen, M. S. (2014). MNE software for processing MEG and EEG data. *NeuroImage*, *86*, 446–460. https://doi.org/10.1016/j.neuroimage.2013.10.027
- Graziano, M. S. A., Reiss, L. A. J., & Gross, C. G. (1999). A neuronal representation of the location of nearby sounds. *Nature*, *397*(6718), 428–430. https://doi.org/10.1038/17115
- Graziano, M. S. A., Yap, G. S., & Gross, C. G. (1994). Coding of visual space by premotor neurons. *Science*, *266*(5187), 1054–1057. https://doi.org/10.1126/science.7973661
- Groh, J. M., Kelly, K. A., & Underhill, A. M. (2003). A Monotonic Code for Sound Azimuth in Primate Inferior Colliculus. *Journal of Cognitive Neuroscience*, *15*(8), 1217–1231. https://doi.org/10.1162/089892903322598166
- 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
- Grothe, B., Pecka, M., & McAlpine, D. (2010). Mechanisms of Sound Localization in Mammals. *Physiological Reviews*, *90*(3), 983–1012. https://doi.org/10.1152/physrev.00026.2009
- Guggenmos, M., Sterzer, P., & Cichy, R. M. (2017). Representational similarity analysis for MEEG: a comprehensive comparison of dissimilarity measures. *BioRxiv Preprint*, 0–35. https://doi.org/http://dx.doi.org/10.1101/172619
- Guzman-Martinez, E., Ortega, L., Grabowecky, M., Mossbridge, J., & Suzuki, S. (2012). Interactive coding of visual spatial frequency and auditory amplitudemodulation rate. *Current Biology*, 22(5), 383–388. https://doi.org/10.1016/j.cub.2012.01.004

Hancock, K. E., & Delgutte, B. (2004). A physiologically based model of interaural

time difference discrimination. *The Journal of Neuroscience : The Official Journal of the Society for Neuroscience, 24*(32), 7110–7117. https://doi.org/10.1523/JNEUROSCI.0762-04.2004

- Hansen, L. K., Larsen, J., Nielsen, F. Å., Strother, S. C., Rostrup, E., Savoy, R., ... Paulson, O. B. (1999). Generalizable patterns in neuroimaging: How many principal components? *NeuroImage*, 9(5), 534–544. https://doi.org/10.1006/nimg.1998.0425
- Harper, N. S., & McAlpine, D. (2004). Optimal neural population coding of an auditory spatial cue. *Nature*, *430*(7000), 682–686. https://doi.org/10.1038/nature02768
- Hastie, T., Tibshirani, R., & Friedman, J. (2009). *The Elements of Statistical Learning*. New York, NY: Springer New York. https://doi.org/10.1007/978-0-387-84858-7
- Haufe, S., Meinecke, F., Görgen, K., Dähne, S., Haynes, J. D., Blankertz, B., & Bießmann, F. (2014). On the interpretation of weight vectors of linear models in multivariate neuroimaging. *NeuroImage*, 87, 96–110. https://doi.org/10.1016/j.neuroimage.2013.10.067
- Haynes, J. D. (2015). A Primer on Pattern-Based Approaches to fMRI: Principles, Pitfalls, and Perspectives. *Neuron*, *87*(2), 257–270. https://doi.org/10.1016/j.neuron.2015.05.025
- Hebart, M. N., Görgen, K., & Haynes, J.-D. (2015). The Decoding Toolbox (TDT): a versatile software package for multivariate analyses of functional imaging data. *Frontiers in Neuroinformatics*, 8(January), 1–18. https://doi.org/10.3389/fninf.2014.00088
- Held, R. (1965). Plasticity in sensory-motor systems. *Scientific American*, *213*(5), 84–94. https://doi.org/10.1038/scientificamerican1165-84
- Holmes, N. P. (2007). The law of inverse effectiveness in neurons and behaviour: Multisensory integration versus normal variability. *Neuropsychologia*, 45(14), 3340–3345. https://doi.org/10.1016/j.neuropsychologia.2007.05.025
- Holmes, N. P. (2009). The principle of inverse effectiveness in multisensory integration: Some statistical considerations. *Brain Topography*, *21*(3–4), 168–176. https://doi.org/10.1007/s10548-009-0097-2
- 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
- Hupé, J.-M., Joffo, L.-M., & Pressnitzer, D. (2008). Bistability for audiovisual stimuli: Perceptual decision is modality specific. *Journal of Vision*, *8*(7), 1. https://doi.org/10.1167/8.7.1
- Itatani, N., & Klump, G. M. (2011). Neural Correlates of Auditory Streaming of Harmonic Complex Sounds With Different Phase Relations in the Songbird Forebrain. *Journal of Neurophysiology*, *105*(1), 188–199. https://doi.org/10.1152/jn.00496.2010

- Jeffress, L. A. (1948). A place theory of sound localization. *Journal of Comparative* and Physiological Psychology, 41(1), 35–39. https://doi.org/10.1037/h0061495
- Johnson, R. W. (2001). An Introduction to the Bootstrap. Teaching Statistics (Vol. 23). Boston, MA: Springer US. https://doi.org/10.1111/1467-9639.00050
- Joris, P., & Yin, T. C. T. (2007). A matter of time: internal delays in binaural processing. *Trends in Neurosciences*, *30*(2), 70–78. https://doi.org/10.1016/j.tins.2006.12.004
- Kass, R. E., & Raftery, A. E. (1995). Bayes factors. *Journal of the American Statistical Association*, *90*(430), 773–795. https://doi.org/10.1080/01621459.1995.10476572
- Kawecki, T. J., & Ebert, D. (2004). Conceptual issues in local adaptation. *Ecology Letters*, 7(12), 1225–1241. https://doi.org/10.1111/j.1461-0248.2004.00684.x
- Kayser, C., Petkov, C. I., Augath, M., & Logothetis, N. K. (2005). Integration of touch and sound in auditory cortex. *Neuron*, *48*(2), 373–384. https://doi.org/10.1016/j.neuron.2005.09.018
- Kayser, C., Petkov, C. I., & Logothetis, N. K. (2008). Visual modulation of neurons in auditory cortex. *Cerebral Cortex*, 18(7), 1560–1574. https://doi.org/10.1093/cercor/bhm187
- Kingdom, F. A. A., & Prins, N. (2010). *Psychophysics: a practical introduction*. Academic Press.
- Kingdom, F. A. A., & Prins, N. (2016). *Psychophysics (Second Edition)*. (F. A. A. Kingdom & N. Prins, Eds.). San Diego: Academic Press.
- Kleiner, M., Brainard, D. H., & Pelli, D. G. (2007). What's new in Psychtoolbox-3? In *Perception* (Vol. 36, p. 14). Alezzo. https://doi.org/10.1068/v070821
- Klink, P. C., van Wezel, R. J. A., & van Ee, R. (2012). United we sense, divided we fail: Context-driven perception of ambiguous visual stimuli. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 367(1591), 932–941. https://doi.org/10.1098/rstb.2011.0358
- Knudsen, E. I., & Knudsen, P. F. (1989). Vision calibrates sound localization in developing barn owls. The Journal of Neuroscience: The Official Journal of the Society for Neuroscience, 9(9), 3306–3313. https://doi.org/citeulike-articleid:409336
- Koch, C., & Tsuchiya, N. (2007). Attention and consciousness: two distinct brain processes. *Trends in Cognitive Sciences*, *11*(1), 16–22. https://doi.org/10.1016/j.tics.2006.10.012
- Koch, C., & Tsuchiya, N. (2012). Attention and consciousness: Related yet different. *Trends in Cognitive Sciences*, 16(2), 103–105. https://doi.org/10.1016/j.tics.2011.11.012
- Kong, L., Michalka, S. W., Rosen, M. L., Sheremata, S. L., Swisher, J. D., Shinn-Cunningham, B. G., & Somers, D. C. (2014). Auditory spatial attention
representations in the human cerebral cortex. *Cerebral Cortex*, 24(3), 773–784. https://doi.org/10.1093/cercor/bhs359

- Körding, K. P., Beierholm, U., Ma, W. J., Quartz, S., Tenenbaum, J. B., & Shams, L. (2007). Causal inference in multisensory perception. *PLoS ONE*, 2(9), e943. https://doi.org/10.1371/journal.pone.0000943
- Kriegeskorte, N. (2008). Representational similarity analysis connecting the branches of systems neuroscience. *Frontiers in Systems Neuroscience*, 2, 4. https://doi.org/10.3389/neuro.06.004.2008
- Kriegeskorte, N., & Kievit, R. A. (2013). Representational geometry: Integrating cognition, computation, and the brain. *Trends in Cognitive Sciences*, 17(8), 401– 412. https://doi.org/10.1016/j.tics.2013.06.007
- Krugliak, A., & Noppeney, U. (2016). Synaesthetic interactions across vision and audition. *Neuropsychologia*, *88*, 65–73. https://doi.org/10.1016/j.neuropsychologia.2015.09.027
- Krumbholz, K., Schönwiesner, M., Rübsamen, R., Zilles, K., Fink, G. R., & Von Cramon, D. Y. (2005). Hierarchical processing of sound location and motion in the human brainstem and planum temporale. *European Journal of Neuroscience*, 21(1), 230–238. https://doi.org/10.1111/j.1460-9568.2004.03836.x
- 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
- 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
- Ledoit, O., & Wolf, M. (2004). A well-conditioned estimator for large-dimensional covariance matrices. *Journal of Multivariate Analysis*, *88*(2), 365–411. https://doi.org/10.1016/S0047-259X(03)00096-4
- Lee, H., & Noppeney, U. (2011). Physical and Perceptual Factors Shape the Neural Mechanisms That Integrate Audiovisual Signals in Speech Comprehension. *Journal of Neuroscience*, *31*(31), 11338–11350. https://doi.org/10.1523/JNEUROSCI.6510-10.2011
- 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
- Lehmann, C., Herdener, M., Esposito, F., Hubl, D., di Salle, F., Scheffler, K., ... Seifritz, E. (2006). Differential patterns of multisensory interactions in core and belt areas of human auditory cortex. *NeuroImage*, 31(1), 294–300. https://doi.org/10.1016/j.neuroimage.2005.12.038

Leinonen, L., Hyvarinen, J., & Sovijarvi, A. R. (1980). Functional properties of

neurons in the temporo-parietal association cortex of awake monkey. *Experimental Brain Research*, *39*(2), 203–215. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/6772459

- Lewald, J. (2002). Rapid adaptation to auditory-visual spatial disparity. *Learning and Memory*, *9*(5), 268–278. https://doi.org/10.1101/lm.51402
- Lewald, J., Foltys, H., & Töpper, R. (2002). Role of the posterior parietal cortex in spatial hearing. *The Journal of Neuroscience : The Official Journal of the Society for Neuroscience*, 22(3), RC207. https://doi.org/Rc207
- Lewis, R., & Noppeney, U. (2010). Audiovisual Synchrony Improves Motion Discrimination via Enhanced Connectivity between Early Visual and Auditory Areas. *Journal of Neuroscience*, 30(37), 12329–12339. https://doi.org/10.1523/JNEUROSCI.5745-09.2010
- Lunghi, C., & Alais, D. (2013). Touch Interacts with Vision during Binocular Rivalry with a Tight Orientation Tuning. *PLoS ONE*, *8*(3), e58754. https://doi.org/10.1371/journal.pone.0058754
- Lunghi, C., Binda, P., & Morrone, M. C. (2010, March 23). Touch disambiguates rivalrous perception at early stages of visual analysis. *Current Biology*. https://doi.org/10.1016/j.cub.2009.12.015
- Lunghi, C., Morrone, M. C., & Alais, D. (2014). Auditory and Tactile Signals Combine to Influence Vision during Binocular Rivalry. *The Journal of Neuroscience*, *34*(3), 784–792. https://doi.org/10.1523/JNEUROSCI.2732-13.2014
- Lunghi, C., Verde, L. Lo, & Alais, D. (2017). Touch accelerates visual awareness. *I-Perception*, *8*(1), 204166951668698. https://doi.org/10.1177/2041669516686986
- Ma, W. J., Zhou, X., Ross, L. A., Foxe, J. J., & Parra, L. C. (2009). Lip-reading aids word recognition most in moderate noise: A Bayesian explanation using highdimensional feature space. *PLoS ONE*, *4*(3), e4638. https://doi.org/10.1371/journal.pone.0004638
- Macaluso, E., & Driver, J. (2003). Multimodal spatial representations in the human parietal cortex: Evidence from functional imaging. *Advances in Neurology*, *93*, 219–233. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/12894411
- 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
- Macaluso, E., Frith, C. D., & Driver, J. (2000). Modulation of human visual cortex by crossmodal spatial attention. *Science*, *289*(5482), 1206–1208. https://doi.org/10.1126/science.289.5482.1206
- Maiworm, M., Bellantoni, M., Spence, C., & Röder, B. (2012). When emotional valence modulates audiovisual integration. *Attention, Perception, and Psychophysics*, 74(6), 1302–1311. https://doi.org/10.3758/s13414-012-0310-3
- Maris, E., & Oostenveld, R. (2007). Nonparametric statistical testing of EEG- and MEG-data. *Journal of Neuroscience Methods*, 164(1), 177–190.

https://doi.org/10.1016/j.jneumeth.2007.03.024

- Martuzzi, R., Murray, M. M., Michel, C. M., Thiran, J. P., Maeder, P. P., Clarke, S., & Meuli, R. A. (2007). Multisensory interactions within human primary cortices revealed by BOLD dynamics. *Cerebral Cortex*, 17(7), 1672–1679. https://doi.org/10.1093/cercor/bhl077
- 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), 7–7. https://doi.org/10.1167/8.11.7
- McAlpine, D. (2005). Creating a sense of auditory space. *Journal of Physiology*, 566(1), 21–28. https://doi.org/10.1113/jphysiol.2005.083113
- McAlpine, D., Jiang, D., & Palmer, A. R. (2001). A neural code for low-frequency sound localization in mammals. *Nature Neuroscience*, *4*(4), 396–401. https://doi.org/10.1038/86049
- Mendonca, C., Escher, A., van de Par, S., & Colonius, H. (2015). Predicting auditory space calibration from recent multisensory experience. *Experimental Brain Research*, 233(7), 1983–1991. https://doi.org/10.1007/s00221-015-4259-z
- Meredith, M. A., & Allman, B. L. (2009). Subthreshold multisensory processing in cat auditory cortex. *NeuroReport*, 20(2), 126–131. https://doi.org/10.1097/WNR.0b013e32831d7bb6
- Meredith, M. A., & Stein, B. E. (1983). Interactions among converging sensory inputs in the superior colliculus. *Science*, *221*(4608), 389–391. https://doi.org/10.1126/science.6867718
- Meredith, M. A., & Stein, B. E. (1986). Visual, auditory, and somatosensory convergence on cells in superior colliculus results in multisensory integration. *Journal of Neurophysiology*, 56(3), 640–662. https://doi.org/10.1152/jn.1986.56.3.640
- Michalka, S. W., Rosen, M. L., Kong, L., Shinn-Cunningham, B. G., & Somers, D. C. (2016). Auditory Spatial Coding Flexibly Recruits Anterior, but Not Posterior, Visuotopic Parietal Cortex. *Cerebral Cortex*, 26(3), 1302–1308. https://doi.org/10.1093/cercor/bhv303
- Middlebrooks, J. C., & Knudsen, E. I. (1984). A neural code for auditory space in the cat's superior colliculus. *J Neurosci*, *4*(10), 2621–2634. https://doi.org/10.1523/JNEUROSCI.04-10-02621.1984
- Mihalik, A. (2017, December). *The neural basis of audio-visual integration and adaptation*. University of Birmingham. Retrieved from http://etheses.bham.ac.uk/7692/
- Mills, A. W. (1958). On the Minimum Audible Angle. *The Journal of the Acoustical Society of America*, *30*(4), 237–246. https://doi.org/10.1121/1.1909553
- Molholm, S., Ritter, W., Murray, M. M., Javitt, D. C., Schroeder, C. E., & Foxe, J. J. (2002). Multisensory auditory-visual interactions during early sensory processing in humans: A high-density electrical mapping study. *Cognitive Brain Research*,

14(1), 115–128. https://doi.org/10.1016/S0926-6410(02)00066-6

- Moore, B. (2013). An Introduction to the Psychology of Hearing. Leiden, The Netherlands: Brill. https://doi.org/https://doi.org/
- Morgan, M. L., DeAngelis, G. C., & Angelaki, D. E. (2008). Multisensory Integration in Macaque Visual Cortex Depends on Cue Reliability. *Neuron*, 59(4), 662–673. https://doi.org/10.1016/j.neuron.2008.06.024
- Morrell, F. (1972). Visual system's view of acoustic space. *Nature*, 238(5358), 44–46. https://doi.org/10.1038/238044a0
- Morrone, M. C., & Lunghi, C. (2013). Early interaction between vision and touch during binocular rivalry. *Multisensory Research*, *26*(3), 291–306. https://doi.org/10.1163/22134808-00002411
- Mozolic, J. L., Hugenschmidt, C. E., Peiffer, A. M., & Laurienti, P. J. (2008). Modalityspecific selective attention attenuates multisensory integration. *Experimental Brain Research*, *184*(1), 39–52. https://doi.org/10.1007/s00221-007-1080-3
- 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., 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
- Murray, M. M., Molholm, S., Michel, C. M., Heslenfeld, D. J., Ritter, W., Javitt, D. C., ... Foxe, J. J. (2005). Grabbing your ear: Rapid auditory-somatosensory multisensory interactions in low-level sensory cortices are not constrained by stimulus alignment. *Cerebral Cortex*, 15(7), 963–974. https://doi.org/10.1093/cercor/bhh197
- Nagy, A., Eördegh, G., Paróczy, Z., Márkus, Z., & Benedek, G. (2006). Multisensory integration in the basal ganglia. *European Journal of Neuroscience*, *24*(3), 917–924. https://doi.org/10.1111/j.1460-9568.2006.04942.x
- Nardo, D., Santangelo, V., & Macaluso, E. (2014). Spatial orienting in complex audiovisual environments. *Human Brain Mapping*, *35*(4), 1597–1614. https://doi.org/10.1002/hbm.22276
- Naselaris, T., Kay, K. N., Nishimoto, S., & Gallant, J. L. (2011). Encoding and decoding in fMRI. *NeuroImage*, 56(2), 400–410. https://doi.org/10.1016/j.neuroimage.2010.07.073
- Nichols, T. E., & Holmes, A. (2003). Nonparametric Permutation Tests for Functional Neuroimaging. In *Human Brain Function: Second Edition* (Vol. 15, pp. 887–910). https://doi.org/10.1016/B978-012264841-0/50048-2
- Nichols, T. E., & Holmes, A. P. (2001). Nonparametric Permutation Tests for {PET} functional Neuroimaging Experiments: A Primer with examples. *Human Brain Mapping*, 15(1), 1–25. https://doi.org/10.1002/hbm.1058

- Nili, H., Wingfield, C., Walther, A., Su, L., Marslen-Wilson, W., & Kriegeskorte, N. (2014). A Toolbox for Representational Similarity Analysis. *PLoS Computational Biology*, *10*(4), e1003553. https://doi.org/10.1371/journal.pcbi.1003553
- Noppeney, U. (2012). Characterization of multisensory integration with fMRI. In *The neural* bases of multisensory processes. (pp. 1–22). https://doi.org/doi:10.1201/b11092-17
- Olivers, C. N. L., & Van der Burg, E. (2008). Bleeping you out of the blink: Sound saves vision from oblivion. *Brain Research*, *1242*(2000), 191–199. https://doi.org/10.1016/j.brainres.2008.01.070
- Onat, S., Libertus, K., & König, P. (2007). Integrating audiovisual information for the control of overt attention. *Journal of Vision*, 7(10), 11. https://doi.org/10.1167/7.10.11
- 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, 156869. https://doi.org/10.1155/2011/156869
- Ortiz-Rios, M., Azevedo, F. A. C., Kuśmierek, P., Balla, D. Z., Munk, M. H., Keliris, G. A., ... Rauschecker, J. P. (2017). Widespread and Opponent fMRI Signals Represent Sound Location in Macaque Auditory Cortex. *Neuron*, *93*(4), 971–983.e4. https://doi.org/10.1016/j.neuron.2017.01.013
- Padberg, J., Seltzer, B., & Cusick, C. G. (2003). Architectonics and Cortical Connections of the Upper Bank of the Superior Temporal Sulcus in the Rhesus Monkey: An Analysis in the Tangential Plane. *Journal of Comparative Neurology*, 467(3), 418–434. https://doi.org/10.1002/cne.10932
- 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
- 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), 969. https://doi.org/10.3389/fpsyg.2014.00969
- Porter, J. R., Semenov, M. A., Hulme, M., Harrison, P., & Arnell, N. (1999). Climate variability and crop yields in Europe [8]. *Nature*, *400*(6746), 724. https://doi.org/10.1038/23385

Radeau, M., & Bertelson, P. (1974). The after-effects of ventriloquism. The Quarterly

Journal of Experimental Psychology, *26*(1), 63–71. https://doi.org/10.1080/14640747408400388

- Radeau, M., & Bertelson, P. (1976). The effect of a textured visual field on modality dominance in a ventriloquism situation. *Perception & Psychophysics*, 20(4), 227–235. https://doi.org/10.3758/BF03199448
- 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
- Ramkumar, P., Jas, M., Pannasch, S., Hari, R., & Parkkonen, L. (2013). Feature-Specific Information Processing Precedes Concerted Activation in Human Visual Cortex. *Journal of Neuroscience*, 33(18), 7691–7699. https://doi.org/10.1523/JNEUROSCI.3905-12.2013
- Rauschecker, J. P., & Tian, B. (2000). Mechanisms and streams for processing of "what" and "where" in auditory cortex. *Proceedings of the National Academy of Sciences*, 97(22), 11800–11806. https://doi.org/10.1073/pnas.97.22.11800
- Rayleigh, Lord. (1907). On our perception of sound direction. *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science, 13*(74), 214–232. https://doi.org/10.1080/14786440709463595
- Recanzone, G. H. (1998). Rapidly induced auditory plasticity: The ventriloquism aftereffect. *Proceedings of the National Academy of Sciences*, *95*(3), 869–875. https://doi.org/10.1073/pnas.95.3.869
- Recanzone, G. H. (2009). Interactions of auditory and visual stimuli in space and time. *Hearing Research*, 258(1–2), 89–99. https://doi.org/10.1016/j.heares.2009.04.009
- Recanzone, G. H. (2011). The biological basis of audition. *Wiley Interdisciplinary Reviews: Cognitive Science*, 2(4), 408–418. https://doi.org/10.1002/wcs.118
- Reyes, A., Rubel, E., & Spain, W. (1996). In vitro analysis of optimal stimuli for phase-locking and time-delayed modulation of firing in avian nucleus laminaris neurons. *The Journal of Neuroscience*, 16(3), 993–1007. https://doi.org/10.1523/JNEUROSCI.16-03-00993.1996
- Rigoux, L., Stephan, K. E., Friston, K. J., & Daunizeau, J. (2014). Bayesian model selection for group studies - Revisited. *NeuroImage*, 84, 971–985. https://doi.org/10.1016/j.neuroimage.2013.08.065
- Rock, I., & Victor, J. (1964). Vision and touch: An experimentally created conflict between the two senses. *Science*, *143*(3606), 594–596. https://doi.org/10.1126/science.143.3606.594
- 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
- Romanski, L. M. (2007). Representation and integration of auditory and visual stimuli in the primate ventral lateral prefrontal cortex. *Cerebral Cortex*, *17*(SUPPL. 1), 61–69. https://doi.org/10.1093/cercor/bhm099
- Salminen, N. H., May, P. J. C., Alku, P., & Tiitinen, H. (2009). A population rate code of auditory space in the human cortex. *PLoS ONE*, *4*(10), e7600. https://doi.org/10.1371/journal.pone.0007600
- Salminen, N. H., Tiitinen, H., Yrttiaho, S., & May, P. J. C. (2010). The neural code for interaural time difference in human auditory cortex. *The Journal of the Acoustical Society of America*, 127(2), EL60-EL65. https://doi.org/10.1121/1.3290744
- Santangelo, V., Belardinelli, M. O., Spence, C., & Macaluso, E. (2009). Interactions between voluntary and stimulus-driven spatial attention mechanisms across sensory modalities. *Journal of Cognitive Neuroscience*, 21(12), 2384–2397. https://doi.org/10.1162/jocn.2008.21178
- Santangelo, V., & Macaluso, E. (2012). Spatial attention and audiovisual processing. In B. E. Stein (Ed.), *The New Handbook of Multisensory Processing* (pp. 359–370). MIT Press.
- Santangelo, V., & Spence, C. (2008). Is the exogenous orienting of spatial attention truly automatic? Evidence from unimodal and multisensory studies. *Consciousness and Cognition*, 17(3), 989–1015. https://doi.org/10.1016/j.concog.2008.02.006
- Sato, Y., Toyoizumi, T., & Aihara, K. (2007). Bayesian inference explains perception of unity and ventriloquism aftereffect: Identification of common sources of audiovisual stimuli. *Neural Computation*, 19(12), 3335–3355. https://doi.org/10.1162/neco.2007.19.12.3335
- Schindler, A., Herdener, M., & Bartels, A. (2013). Coding of Melodic Gestalt in Human Auditory Cortex. *Cerebral Cortex*, 23(12), 2987–2993. https://doi.org/10.1093/cercor/bhs289
- Schmahmann, J. D., & Pandya, D. N. (1991). Projections to the basis pontis from the superior temporal sulcus and superior temporal region in the rhesus monkey. *Journal of Comparative Neurology*, 308(2), 224–248. https://doi.org/10.1002/cne.903080209
- 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
- Schrouff, J., Rosa, M. J., Rondina, J. M., Marquand, A. F., Chu, C., Ashburner, J., ... Mourão-Miranda, J. (2013). PRoNTo: Pattern recognition for neuroimaging toolbox. *Neuroinformatics*, *11*(3), 319–337. https://doi.org/10.1007/s12021-013-9178-1

- Shams, L., Kamitani, Y., & Shimojo, S. (2000). Illusions: What you see is what you hear. *Nature*, *408*(6814), 788. https://doi.org/10.1038/35048669
- Shams, L., 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
- Shepard, R. N. (1980). Multidimensional scaling, tree-fitting, and clustering. *Science*, *210*(4468), 390–398. https://doi.org/10.1126/science.210.4468.390
- 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
- Soto-Faraco, S., & Spence, C. (2002). Modality-specific auditory and visual temporal processing deficits. *The Quarterly Journal of Experimental Psychology Section A*, 55(1), 23–40. https://doi.org/10.1080/02724980143000136
- Spence, C. (2010). Crossmodal spatial attention. *Annals of the New York Academy* of Sciences, 1191, 182–200. https://doi.org/10.1111/j.1749-6632.2010.05440.x
- Spence, C. (2013). Just how important is spatial coincidence to multisensory integration? Evaluating the spatial rule. *Annals of the New York Academy of Sciences*, 1296(1), 31–49. https://doi.org/10.1111/nyas.12121
- Spinelli, D. N., Starr, A., & Barrett, T. W. (1968). Auditory specificity in unit recordings from cat's visual cortex. *Experimental Neurology*, 22(1), 75–84. https://doi.org/10.1016/0014-4886(68)90020-4
- Stanford, T. R. (2005). Evaluating the Operations Underlying Multisensory Integration in the Cat Superior Colliculus. *Journal of Neuroscience*, *25*(28), 6499–6508. https://doi.org/10.1523/JNEUROSCI.5095-04.2005
- Stecker, G. C., Harrington, I. A., & Middlebrooks, J. C. (2005). Location coding by opponent neural populations in the auditory cortex. *PLoS Biology*, *3*(3), 0520–0528. https://doi.org/10.1371/journal.pbio.0030078
- Stein, B. E. (1978). Development and organization of multimodal representation in cat superior colliculus. *Federation Proceedings*, 37(9), 2240–2245. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/658465
- Stein, B. E., & Arigbede, M. O. (1972). Unimodal and multimodal response properties of neurons in the cat's superior colliculus. *Experimental Neurology*, 36(1), 179– 196. https://doi.org/10.1016/0014-4886(72)90145-8
- Stein, B. E., Magalhaes-Castro, B., & Kruger, L. (1975). Superior colliculus: visuotopic-somatotopic overlap. *Science (New York, N.Y.)*, 189(4198), 224–226. https://doi.org/10.1126/science.1094540
- Stein, B. E., & Meredith, M. A. (1993). *The Merging of the Senses. Cambridge, MA.: A Bradford Book.* Cambridge, MA, US: The MIT Press.
- Stein, B. E., Stanford, T. R., Ramachandran, R., Perrault, T. J., & Rowland, B. A. (2009). Challenges in quantifying multisensory integration: Alternative criteria,

models, and inverse effectiveness. *Experimental Brain Research*, 198(2–3), 113–126. https://doi.org/10.1007/s00221-009-1880-8

- 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
- Stern, R. M., & Trahiotis, C. (1995). Models of binaural interaction. In *Handbook of perception and cognition* (Vol. 6, pp. 347–386).
- Stern, R. M., Zeiberg, A. S., & Trahiotis, C. (1988). Lateralization of complex binaural stimuli: a weighted-image model. J Acoust Soc Am, 84(1), 156–165. https://doi.org/10.1121/1.396982
- Sugihara, T., Diltz, M. D., Averbeck, B. B., & Romanski, L. M. (2006). Integration of Auditory and Visual Communication Information in the Primate Ventrolateral Prefrontal Cortex. *Journal of Neuroscience*, 26(43), 11138–11147. https://doi.org/10.1523/JNEUROSCI.3550-06.2006
- Talsma, D., Doty, T. J., & Woldorff, M. G. (2007). Selective attention and audiovisual integration: Is attending to both modalities a prerequisite for early integration? *Cerebral Cortex*, *17*(3), 679–690. https://doi.org/10.1093/cercor/bhk016
- Talsma, D., Senkowski, D., Soto-Faraco, S., & Woldorff, M. G. (2010). The multifaceted interplay between attention and multisensory integration. *Trends in Cognitive Sciences*, *14*(9), 400–410. https://doi.org/10.1016/j.tics.2010.06.008
- Talsma, D., & Woldorff, M. G. (2005). Selective attention and multisensory integration: Multiple phases of effects on the evoked brain activity. *Journal of Cognitive Neuroscience*, *17*(7), 1098–1114. https://doi.org/10.1162/0898929054475172
- Teder-Sälejärvi, W. A., Di Russo, F., McDonald, J. J., & Hillyard, S. A. (2005). Effects of spatial congruity on audio-visual multimodal integration. *Journal of Cognitive Neuroscience*, *17*(9), 1396–1409. https://doi.org/10.1162/0898929054985383
- Teder-Sälejärvi, W. A., McDonald, J. J., Di Russo, F., & Hillyard, S. A. (2002). An analysis of audio-visual crossmodal integration by means of event-related potential (ERP) recordings. *Cognitive Brain Research*, *14*(1), 106–114. https://doi.org/10.1016/S0926-6410(02)00065-4
- Thompson, S. K., Von Kriegstein, K., Deane-Pratt, A., Marquardt, T., Deichmann, R., Griffiths, T. D., & McAlpine, D. (2006). Representation of interaural time delay in the human auditory midbrain. *Nature Neuroscience*, 9(9), 1096–1098. https://doi.org/10.1038/nn1755

Tollin, D. J., & Yin, T. C. T. (2002). The Coding of Spatial Location by Single Units in

the Lateral Superior Olive of the Cat. I. Spatial Receptive Fields in Azimuth. *The Journal of Neuroscience*, 22(4), 1454–1467. https://doi.org/10.1523/JNEUROSCI.22-04-01454.2002

- Tong, F. (2003). Cognitive neuroscience: Primary visual cortex and visual awareness. *Nature Reviews Neuroscience*, *4*(3), 219–229. https://doi.org/10.1038/nrn1055
- Tononi, G., & Edelman, G. M. (1998). Consciousness and complexity. *Science*, *282*(5395), 1846–1851. https://doi.org/10.1126/science.282.5395.1846
- Trahiotis, C., Bernstein, L. R., & Akeroyd, M. A. (2001). Manipulating the "straightness" and "curvature" of patterns of interaural cross correlation affects listeners' sensitivity to changes in interaural delay. *The Journal of the Acoustical Society of America*, 109(1), 321–330. https://doi.org/10.1121/1.1327579
- Trahiotis, C., & Stern, R. M. (1989). Lateralization of bands of noise: Effects of bandwidth and differences of interaural time and phase. *The Journal of the Acoustical Society of America*, 86(4), 1285–1293. https://doi.org/10.1121/1.398743
- Trapeau, R., & Schönwiesner, M. (2015). Adaptation to shifted interaural time differences changes encoding of sound location in human auditory cortex. *NeuroImage*, *118*, 26–38. https://doi.org/10.1016/j.neuroimage.2015.06.006
- Tsuchiya, N., & Koch, C. (2005). Continuous flash suppression reduces negative afterimages. *Nature Neuroscience*, *8*(8), 1096–1101. https://doi.org/10.1038/nn1500
- Urbantschitsch, V. (1880). Über den Einfluss einer Sinneserregung auf die übrigen Sinnesempfindungen. Archiv f d Gesch Physiol, 42, 155.
- Van Atteveldt, N., Formisano, E., Goebel, R., & Blomert, L. (2004). Integration of letters and speech sounds in the human brain. *Neuron*, 43(2), 271–282. https://doi.org/10.1016/j.neuron.2004.06.025
- van Ee, R., van Boxtel, J. J. A., Parker, A. L., & Alais, D. (2009). Multisensory Congruency as a Mechanism for Attentional Control over Perceptual Selection. *Journal of Neuroscience*, 29(37), 11641–11649. https://doi.org/10.1523/JNEUROSCI.0873-09.2009
- Van Essen, D. C., Anderson, C. H., & Felleman, D. J. (1992). Information processing in the primate visual system: An integrated systems perspective. *Science*, 255(5043), 419–423. https://doi.org/10.1126/science.1734518
- 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
- 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

- von Békésy, G. (1930). Zur Theorie des Hörens. Über das Richtungshören bei einer Zeitdifferenz oder Lautstärkenungleichheit der beiderseitigen Schalleinwirkungen. *Physik. Z.*, *31*, 824–835.
- von Helmholtz, H. (1896). Handbuch der physiologischen Optik. Monatshefte für Mathematik und Physik (Vol. 7). Leipzig, Germany: Leopold Voss. https://doi.org/10.1007/BF01708548
- von Kriegstein, K., Griffiths, T. D., Thompson, S. K., & McAlpine, D. (2008). Responses to Interaural Time Delay in Human Cortex. *Journal of Neurophysiology*, *100*(5), 2712–2718. https://doi.org/10.1152/jn.90210.2008
- von Saldern, S., & Noppeney, U. (2013). Sensory and Striatal Areas Integrate Auditory and Visual Signals into Behavioral Benefits during Motion Discrimination. *Journal of Neuroscience*, *33*(20), 8841–8849. https://doi.org/10.1523/JNEUROSCI.3020-12.2013
- Vonderschen, K., & Wagner, H. (2014). Detecting interaural time differences and remodeling their representation. *Trends in Neurosciences*, *37*(5), 289–300. https://doi.org/10.1016/j.tins.2014.03.002
- Vroomen, J., Bertelson, P., & De Gelder, B. (2001). The ventriloquist effect does not depend on the direction of automatic visual attention. *Perception and 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
- Walther, A., Nili, H., Ejaz, N., Alink, A., Kriegeskorte, N., & Diedrichsen, J. (2016). Reliability of dissimilarity measures for multi-voxel pattern analysis. *NeuroImage*, 137, 188–200. https://doi.org/10.1016/j.neuroimage.2015.12.012
- Wandell, B. A., Dumoulin, S. O., & Brewer, A. A. (2007, October). Visual field maps in human cortex. *Neuron*. https://doi.org/10.1016/j.neuron.2007.10.012
- Wang, L., Mruczek, R. E. B., Arcaro, M. J., & Kastner, S. (2015). Probabilistic maps of visual topography in human cortex. *Cerebral Cortex*, 25(10), 3911–3931. https://doi.org/10.1093/cercor/bhu277
- Watanabe, M., Cheng, K., Murayama, Y., Ueno, K., Asamizuya, T., Tanaka, K., & Logothetis, N. (2011). Attention but not awareness modulates the BOLD signal in the human V1 during binocular suppression. *Science*, 334(6057), 829–831. https://doi.org/10.1126/science.1203161
- Watanabe, T., & Flavin, M. (1976). Nucleotide metabolizing enzymes in Chlamydomonas flagella. *Journal of Biological Chemistry*, 251(1), 182–192. https://doi.org/10.2307/271063
- Watkins, S., Shams, L., Tanaka, S., Haynes, J. D., & Rees, G. (2006). Sound alters activity in human V1 in association with illusory visual perception. *NeuroImage*, *31*(3), 1247–1256. https://doi.org/10.1016/j.neuroimage.2006.01.016

- Weeks, R. A., Aziz-Sultan, A., Bushara, K. O., Tian, B., Wessinger, C. M., Dang, N., ... Hallett, M. (1999). A PET study of human auditory spatial processing. *Neuroscience Letters*, 262(3), 155–158. https://doi.org/10.1016/S0304-3940(99)00062-2
- Welch, R. B. (1999). Chapter 15 Meaning, attention, and the "unity assumption" in the intersensory bias of spatial and temporal perceptions. In Advances in Psychology (Vol. 129, pp. 371–387). Amsterdam: North-Holland/Elsevier Science Publishers. https://doi.org/10.1016/S0166-4115(99)80036-3
- Welch, R. B., & Warren, D. H. (1986). Intersensory interactions. In K. . Boff (Ed.), Handbook of perception and human performance: Sensory processes and perception (Vol. I, p. 25.1-25.36). Wiley.
- Werner-Reiss, U., & Groh, J. M. (2008). A Rate Code for Sound Azimuth in Monkey Auditory Cortex: Implications for Human Neuroimaging Studies. *Journal of Neuroscience*, 28(14), 3747–3758. https://doi.org/10.1523/JNEUROSCI.5044-07.2008
- Werner-Reiss, U., Kelly, K. a, Trause, A. S., Underhill, A. M., & Groh, J. M. (2003). Eye position affects activity in primary auditory cortex of primates. *Current Biology : CB*, 13(7), 554–562. https://doi.org/10.1016/S
- 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
- Wichmann, F. A., & Hill, N. J. (2001). The psychometric function: I. Fitting, sampling, and goodness of fit. *Perception & Psychophysics*, *63*(8), 1293–1313. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/11800458
- Wickelgren, B. G. (1971). Superior colliculus: Some receptive field properties of bimodally responsive cells. *Science*, 173(3991), 69–72. https://doi.org/10.1126/science.173.3991.69
- Witten, I. B., & Knudsen, E. I. (2005). Why seeing is believing: Merging auditory and visual worlds. *Neuron*, *48*(3), 489–496. https://doi.org/10.1016/j.neuron.2005.10.020
- Woods, T. M., Lopez, S. E., Long, J. H., Rahman, J. E., & Recanzone, G. H. (2006). Effects of Stimulus Azimuth and Intensity on the Single-Neuron Activity in the Auditory Cortex of the Alert Macaque Monkey. *Journal of Neurophysiology*, 96(6), 3323–3337. https://doi.org/10.1152/jn.00392.2006
- Woods, T. M., & Recanzone, G. H. (2004). Visually induced plasticity of auditory spatial perception in macaques. *Current Biology*, *14*(17), 1559–1564. https://doi.org/10.1016/j.cub.2004.08.059
- Wozny, D. R., Beierholm, U. R., & Shams, L. (2008). Human trimodal perception follows optimal statistical inference. *Journal of Vision*, *8*(3), 24. https://doi.org/10.1167/8.3.24

- Wozny, D. R., Beierholm, U. R., & Shams, L. (2010). Probability matching as a computational strategy used in perception. *PLoS Computational Biology*, *6*(8). https://doi.org/10.1371/journal.pcbi.1000871
- Wozny, D. R., & Shams, L. (2011a). Computational Characterization of Visually Induced Auditory Spatial Adaptation. *Frontiers in Integrative Neuroscience*, 5(November), 75. https://doi.org/10.3389/fnint.2011.00075
- Wozny, D. R., & Shams, L. (2011b). Recalibration of Auditory Space following Milliseconds of Cross-Modal Discrepancy. *Journal of Neuroscience*, 31(12), 4607–4612. https://doi.org/10.1523/JNEUROSCI.6079-10.2011
- Wyart, V., & Tallon-Baudry, C. (2008). Neural Dissociation between Visual Awareness and Spatial Attention. *Journal of Neuroscience*, *28*(10), 2667–2679. https://doi.org/10.1523/JNEUROSCI.4748-07.2008
- 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
- Zaidel, A., Ma, W. J., & Angelaki, D. E. (2013). Supervised calibration relies on the multisensory percept. *Neuron*, *80*(6), 1544–1557. https://doi.org/10.1016/j.neuron.2013.09.026
- Zaidel, A., Turner, A. H., & Angelaki, D. E. (2011). Multisensory Calibration Is Independent of Cue Reliability. *Journal of Neuroscience*, *31*(39), 13949–13962. https://doi.org/10.1523/JNEUROSCI.2732-11.2011
- 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
- Zierul, B., Röder, B., Tempelmann, C., Bruns, P., & Noesselt, T. (2017). The role of auditory cortex in the spatial ventriloquism aftereffect. *NeuroImage*, *162*, 257–268. https://doi.org/10.1016/j.neuroimage.2017.09.002