THE NEURAL BASIS OF AUDIO-VISUAL INTEGRATION AND ADAPTATION

by Dr. AGOSTON MIHALIK

A thesis submitted to the University of Birmingham for the degree of

DOCTOR OF PHILOSOPHY

School of Psychology

College of Life and Environmental Sciences

University of Birmingham

March 2017

ABSTRACT

The brain integrates or segregates audio-visual signals effortlessly in everyday life. In order to do so, it needs to infer the causal structure by which the signals were generated. Although behavioural studies extensively characterized causal inference in audio-visual perception, the neural mechanisms are barely explored. The current thesis sheds light on these neural processes and demonstrates how the brain adapts to dynamic as well as long-term changes in the environmental statistics of audio-visual signals. In *Chapter 1*, I introduce the causal inference problem and demonstrate how spatial audio-visual signals are integrated at the behavioural as well as neural level. In *Chapter 2*, I describe methodological foundations for the following empirical chapters. In *Chapter 3*, I present the neural mechanisms of explicit causal inference and the representations of audio-visual space along the human cortical hierarchy. *Chapter 4* reveals that the brain is able to use recent past to adapt to the dynamically changing environment. In *Chapter 5*, I discuss the neural substrates of encoding auditory space and its adaptive changes in response to spatially conflicting visual signals. Finally, in *Chapter 6*, I summarize the findings of the thesis, its contributions to the literature, and I outline directions for future research.

ACKNOWLEDGEMENTS

First and foremost, I would like to thank Uta Noppeney, for the great and constant support throughout the PhD. Her dedication to research is heroic. The long evening meetings will be always remembered… I also feel honoured that our relationship went over the strictly defined research projects and we discussed academic as well as personal questions.

I would like to say a huge thank to Thomas White for his great help with proof reading the thesis and correcting my accidental typos and grammar mistakes. Also, it was great to have Tom as a colleague and a friend.

There are two more people I would like to highlight. The first is Mate Aller, my friend and colleague. I can't count that numbers of hours we spent on discussions about literally everything… He had a huge impact on my views in many different ways. Also, sometimes just as setting an example and showing the way how he approaches questions, research and science.

I would like to say a personal thank to David Meijer, for his friendship and collegiality. I enjoyed his company as a person and as a scientist often times 24/7.

The atmosphere in the Computational Cognitive Neuroimaging Group is just amazing. I loved being part of it! I almost felt like being in a family here. Having personal relationship with everyone and being able to have some words even on the very busy days was a great experience. Those days happened a lot... I would like to thank both the current members (not listed above), namely: Arianna Zuanazzi, Sam Jones, Patrycja Delong, Remi Gau, Steffen Burgers, Johanna Zummer, Ambra Ferrari, Giulio Degano. And I would also like to thank previous members: Joana Leitao and Alexandra Krugliak.

Finally, I would like to thank to my Hungarian and international friends in Birmingham for sharing their life with me during my staying in Birmingham. I also thank to my family and friends at home for their patience and support for many years, so that I can be here now and make a huge step ahead in my life and career.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

CHAPTER 1: GENERAL INTRODUCTION

Human perception is based on five traditional senses: vision, hearing, taste, smell and touch. There are clear differences in our senses, but they convey information from the same entity: our environment. Of course, the processes by which the brain extracts information, differentiates signal from noise and integrates all the signals into a coherent picture are not trivial and have inspired research for many decades. The principles of multisensory integration have been established for more than 20 years (Stein & Meredith, 1993), yet our understanding is limited, and a great deal of discovery is ahead of us to really understand how humans perceive the ever-changing environment through multiple senses.

Integration and segregation of audio-visual signals

To illustrate the fundamental problem of multisensory integration, let us consider two everyday situations. When bombarded concurrently with many signals at busy crossroads, the brain has to decide which sensory information to bind together and which ones to keep separate. It is essential to correctly determine the sources of the signals to avoid confusion and accidents. By contrast, in a quiet two person conversation the signals are limited. Yet, integrating the signals might be challenging, and at the same it could be necessary in order to understand the conversation.

It is important to know that the integration-segregation problem exists in various formulations. The most commonly used include: the binding problem; the correspondence problem; object identity decisions; unity judgement; common source judgement; and causal inference (Figure 1.1; Trommershäuser, Kording, & Landy, 2011).

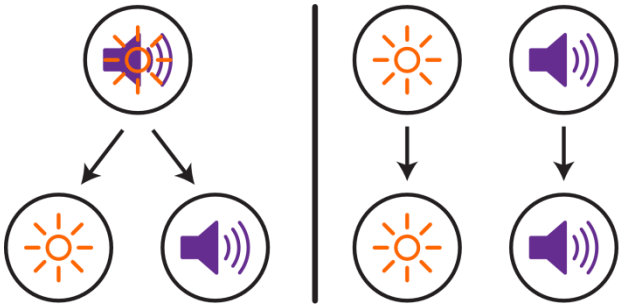


Figure 1.1 Integration and segregation of audio-visual stimuli. Left: Audio-visual signals generated by a single event **Right:** Audio-visual signals generated by multiple events. (Original figure provided by the courtesy of Samuel Jones, adapted for the thesis.)

The current section is divided up to three parts. Firstly, I will describe the various factors the brain utilizes to solve the binding problem. Secondly, I introduce two general mechanisms by which humans process multisensory information. I show how audio-visual integration can be optimal by maximum-likelihood estimation using bottom-up characteristics of sensory signals. Finally, a generative hierarchical Bayesian model is presented that combines bottom-up characteristics of the sensory signals with top-down influences to infer estimates of the sensory signals and solve the binding problem at the same time.

Numerous studies investigated the factors influencing audio-visual integration in the last couple of decades. A series of pioneering studies date back to the 1970s performed by Radeau & Bertelson (Radeau & Bertelson, 1974, 1976, 1977, 1978). After some debate of the nature and importance of certain factors, the general consideration is that both bottom-up sensory correspondences and top-down cognitive mechanism influence audio-visual integration (Bedford, 2001; Recanzone & Sutter, 2008; Talsma, Senkowski, Soto-Faraco, & Woldorff, 2010; Welch, 1999).

The most studied top-down factor is attention. It is considered to influence audio-visual integration at many levels (Santangelo & Macaluso, 2012; Talsma et al., 2010), and attentional effects have been studied in many audio-visual paradigms both in behavioural and neuroimaging studies: the McGurk illusion (Alsius, Navarra, Campbell, & Soto-Faraco, 2005), the ventriloquist effect (Busse, Roberts, Crist, Weissman, & Woldorff, 2005), the double flash illusion (T. S. Andersen, Tiippana, & Sams, 2004) or other AV paradigms (Fairhall & Macaluso, 2009; Johnson & Zatorre, 2005; Mozolic, Hugenschmidt, Peiffer, & Laurienti, 2008; Nardo, Santangelo, & Macaluso, 2014; Santangelo, Olivetti Belardinelli, Spence, & Macaluso, 2009). Perceptual load has also been shown to interfere with audiovisual integration by modulating selective attention (Eramudugolla, Kamke, Soto-Faraco, & Mattingley, 2011). Another focus of cognitive factors concerns the role of participants' belief that the stimuli originate from the same source, which is also called a 'unity assumption' (Bedford, 2001; Warren, Welch, & McCarthy, 1981; Welch & Warren, 1980). It has also been suggested that task instructions have an influence on the unity assumption (Warren et al., 1981). In recent years, more direct relationships have been uncovered between cognitive factors and audio-visual integration. It has been shown that expectations of stimulus characteristics can alter reaction times in an audio-visual integration task (Van Wanrooij, Bremen, & John Van Opstal, 2010). Röder and colleagues demonstrated that emotional (Maiworm, Bellantoni, Spence, & Röder, 2012) and motivational factors (Bruns, Maiworm, & Röder, 2014) can both influence how audio-visual stimuli are bound together.

Three main bottom-up factors help the brain to decide whether audio-visual signals should be integrated or segregated: stimulus spatial and temporal correspondence and their semantic or associative relationship (Chen & Vroomen, 2013; Recanzone, 2009; Slutsky &

Recanzone, 2001; Wallace et al., 2004; Welch, 1999). Integration of AV stimuli degrades with increasing spatial disparity or temporal asynchrony (Slutsky & Recanzone, 2001; Wallace et al., 2004), since the larger the spatio-temporal discrepancy, the less likely the two signals are to originate from a common source. The semantic relationship of audio-visual signals is most studied in speech perception, but can also occur in forms of synesthetic experiences (Calvert, Campbell, & Brammer, 2000; Krugliak & Noppeney, 2016; Laurienti, Kraft, Maldjian, Burdette, & Wallace, 2004; HweeLing Lee & Noppeney, 2011; van Atteveldt, Formisano, Goebel, & Blomert, 2004). The compellingness of the stimuli (Warren et al., 1981; Welch & Warren, 1980) determines how strongly audio-visual signals are associated in a given situation. It includes the number of features redundantly specified by the two signals and so-called historical factors.

So far, we have not addressed an important aspect of audio-visual integration, namely how to process sensory information in an optimal way, and which sensory modality should dominate the integrated percept. Early studies suggested that inter-sensory conflict is resolved based on the *modality precision* or the *modality appropriateness* hypotheses (Welch & Warren, 1980). This hypothesis states that the more precise (e.g. vision versus audition due to its better spatial resolution) or the more appropriate modality (e.g. vision in spatial tasks and audition in temporal tasks) dominates the percept in conflict situations of multisensory signals. These models focus on the qualitative aspect of multisensory integration, and they also do not explain how integration occurs if certain modalities are degraded.

Before we turn to the quantitative models of multisensory integration, I am introducing two general strategies that the brain utilizes to optimize multisensory processing. Multisensory signals can be either complementary or redundant. Multisensory information is complementary if the sensory signals are in different coordinate systems, units or the sensory

signals provide different aspects of the property to be estimated. In this case the optimal strategy appears to maximize the multisensory information via *sensory combination*. Disambiguation is an example for this type of mechanism, when e.g. an accompanying sound creates a motion bounce illusion of moving visual circles. Multisensory information is redundant when the sensory signals refer to the same sensory property using the same coordinate system and unit. In this case the optimal strategy appears to increase the reliability of the sensory estimate via *sensory integration* (Ernst & Bülthoff, 2004).This strategy is mathematically equivalent to the *maximum-likelihood estimation (MLE)* (Ernst & Banks, 2002) The MLE states that the unified sensory percept is weighted by the relative reliabilities of the sensory signals with the more reliable signal to dominate the percept. By definition, the reliability and the variance of a signal are the inverse of each other, therefore the MLE model maximizes the reliability and minimizes the variance of the multisensory estimate at the same time. In their seminal paper, Alais & Burr (Alais & Burr, 2004) showed that audio-visual integration follows MLE and the optimal percept is weighted by the relative reliabilities of the constituent signals. They demonstrated that the more reliable sensory signal always dominates the percept, and it can be either the auditory or the visual stimulus.

MLE has been shown in many studies as an adequate model of bottom-up sensory correspondences, however, a more general approach is needed to account also for top-down influences. *Bayesian decision theory* provides this framework with the introduction of priors (Shams & Beierholm, 2010; Yuille & Bülthoff, 1996). Specifically, the Bayesian model combines sensory representations (*likelihood*) with previous knowledge or the statistics of the environment (*prior*) to make an estimate of the sensory signals (*posterior*) using Bayes' rule. Interestingly, MLE can be considered as a subcase of Bayesian inference when no (uniform) prior is used. It is important to note that the Bayesian decision theory provides a generative

model, in other words, it models explicitly how the signals are generated by events. The first study to apply Bayesian causal inference for audio-visual integration adopted an implicit causal inference model (Shams, Ma, & Beierholm, 2005), but shortly after, a *hierarchical Bayesian model* was introduced that allowed explicit inference not only on the sensory estimates, but on the causal structure of audio-visual signals (Körding et al., 2007; Sato, Toyoizumi, & Aihara, 2007).The hierarchical causal inference proved to be a very fruitful model for audio-visual integration, and different aspects of the model have been further studied or extended in the following years (Beierholm, Quartz, & Shams, 2009; Wozny, Beierholm, & Shams, 2008, 2010; Wozny & Shams, 2011a). For instance, it has been shown that Bayesian priors are encoded independently from likelihoods (Beierholm et al., 2009), and modelling auditory spatial adaptation can be described as a change in the likelihood function (Wozny & Shams, 2011a).

The immediate effect of audio-visual conflict: spatial ventriloquism

The most studied and probably the best known audio-visual integration phenomenon is the *spatial ventriloquist effect* (Bertelson & Aschersleben, 1998; Chen & Vroomen, 2013; Choe, Welch, Gilford, & Juola, 1975; Howard & Templeton, 1966; Radeau & Bertelson, 1976). It refers to the illusory percept by which the apparent location of a sound is perceived toward a visual signal presented simultaneously in a separate location. The phenomenon is named after the ventriloquist situation, when the speech of the performing ventriloquist is mislocalized toward the mouth movements of the puppet. Similarly, the ventriloquist effect happens watching the TV when the speech originating from loudspeakers is mislocalized to the actors' lips. The ventriloquist illusion is a prime example of audio-visual integration, since various top-down and bottom-up factors contributing to a unified multisensory percept (see the previous section) can be studied using this paradigm (Radeau & Bertelson, 1987; Recanzone,

2009; Welch, 1999). Here, I demonstrate some milestones of these studies, and I introduce the discussions that are most relevant for the current thesis.

One of the early discussions in the field concerned the nature of the ventriloquist effect, namely: does the ventriloquist effect reflect a true perceptual phenomenon or is it confounded by decisional bias at some level (Bertelson & Radeau, 1976; Choe et al., 1975)? Choe et al used a very tempting approach applying signal detection theory, and found that participants' unity percept of synchronous vs. asynchronous AV signals was accompanied by a change in their decision criteria, not the perceptual sensitivity between these stimuli (Choe et al., 1975). The study received some critique (Bertelson & Radeau, 1976), and later studies agreed that the ventriloquist effect is based on a perceptual phenomenon (Vroomen & de Gelder, 2004). Different approaches have been provided to minimise or avoid the contamination of AV integration effects by decisional factors. The most common approach is to instruct participants explicitly to ignore the V stimulus during the sound localization task (Bertelson & Radeau, 1981; Howard & Templeton, 1966; Radeau & Bertelson, 1987). Another suggested approach is to use small or undetected AV discrepancies e.g. by applying staircase procedures (Bertelson & Aschersleben, 1998). New experimental designs were proposed that measured the ventriloquist effect indirectly using non-spatial (Driver, 1996), attentional (Spence & Driver, 2000; Vroomen, Bertelson, & de Gelder, 2001a) or auditory motion tasks (Dong, Swindale, & Cynader, 1999). A study using patients with spatial neglect further corroborated the results showing that ventriloquist effect occurs even without awareness of the attracting visual stimulus (Bertelson, Pavani, Ladavas, Vroomen, & de Gelder, 2000).

A similar discussion have been the topic of much interest in the last two decades, questioning whether the ventriloquist effect is an automatic process or can be influenced by

cognitive factors. One of the hot topics regarded the role of attention, where behavioural studies indicated no influence on the illusion (Bertelson, Vroomen, de Gelder, & Driver, 2000; Vroomen, Bertelson, & de Gelder, 2001b). Recent neuroimaging studies provided conflicting evidence using various AV paradigms (Alsius et al., 2005; T. S. Andersen et al., 2004; Busse et al., 2005; Fairhall & Macaluso, 2009; Johnson & Zatorre, 2005; Nardo et al., 2014; Santangelo et al., 2009), however, only one study examined the role of attention in the ventriloquist illusion (Busse et al., 2005). Roder and colleagues investigated the role of motivation and emotion using learning paradigms. They showed that both emotional aversive learning and reward learning can reduce a subsequently measured ventriloquist effect (Bruns et al., 2014; Maiworm et al., 2012). Finally, another aspect of top-down control as prior knowledge has been also demonstrated to influence reaction times in a ventriloquist paradigm (Van Wanrooij et al., 2010). The role and the mechanisms of top-down effects in the ventriloquist illusion hence remains an open question for future studies.

On the other side, there is clear and overwhelming evidence about the role of bottom-up sensory correspondences in the immediate ventriloquist effect. Several studies demonstrated that spatial proximity (Körding et al., 2007; Slutsky & Recanzone, 2001; Wallace et al., 2004) and temporal coincidence (Radeau & Bertelson, 1987; Slutsky & Recanzone, 2001; Thomas, 1941; Wallace et al., 2004) is needed for a strong illusory percept. Early studies also suggested the role of compellingness as a key factor (Warren et al., 1981; Welch & Warren, 1980). In their seminal paper, Alais & Burr (Alais & Burr, 2004) demonstrated that the ventriloquist effect is a near-optimal percept based on reliability weighting and MLE that opened a whole new era of modelling approaches.

It has been shown that Bayesian causal inference provides a more general and generative model for multisensory integration (Bresciani, Dammeier, & Ernst, 2006; Roach,

Heron, & McGraw, 2006; Rowland, Stanford, & Stein, 2007; Shams et al., 2005). Two studies produced evidence in the same year that a hierarchical Bayesian causal inference accounts for the ventriloquist effect (Körding et al., 2007; Sato et al., 2007). Since then, many studies applied the model successfully (Beierholm et al., 2009; Odegaard, Wozny, & Shams, 2015, 2016, Rohe & Noppeney, 2015a, 2015b, 2016).

The aftereffect of audio-visual conflict: visually-induced auditory space adaptation

In addition to the immediate ventriloquist effect, prolonged audio-visual spatial conflict results in the adaptation of the auditory space, called the *ventriloquist aftereffect* (Bertelson, Frissen, Vroomen, & de Gelder, 2006; Canon, 1970, 1971; Chen & Vroomen, 2013; Frissen, Vroomen, de Gelder, & Bertelson, 2003, 2005; Lewald, 2002; Radeau & Bertelson, 1974, 1977; Recanzone, 1998; Wozny & Shams, 2011b). The aftereffect refers to a shift in auditory localization toward the visual signal even when the visual signal is no longer present. The *adaptation* process is also called as *recalibration* or *plasticity* (Held, 1965). The discrepant visual signal is not necessary for auditory space adaptation, in experimental conditions it can occur using ear blocks, ear molds, even altering HRTFs or in natural conditions using electronic hearing devices (Mendonça, 2014).

It is important to note that aftereffects are truly perceptual in contrast to the intricate and multi-faceted immediate effects (Choe et al., 1975). Generally, aftereffects are measured using a design with 3 phases: pre-test, adaptation period and post-test comparing e.g. sound localization performance between pre- and post-test, hence the changes in performance cannot be attributed to cognitive factors or response biases. It could be speculated that response learning can contribute to the adaptation effect, namely, that a tendency to localize in a certain direction during adaptation persists even in the absence of conflicting stimulus (e.g. visual

signal) due to motor learning or response tendency (Choe et al., 1975). Using a different task (e.g. visual detection) during adaptation eliminates this possible confound, that was indeed implemented in most of the classical studies. Interestingly, some recent recalibration experiments interleaved the exposure period with post-test and/or measuring the ventriloquist effect together with the aftereffect raising the concern of contaminating the aftereffect with response biases (Mendonça, Escher, van de Par, & Colonius, 2015; Wozny & Shams, 2011a, 2011b).

Natural conflicts are always present between inter-sensory or more specifically, audio-visual signals (de Gelder & Bertelson, 2003; Ernst & Di Luca, 2011). On one hand, small *temporary conflicts* are perceived between audio-visual signals due to sensory and neural noise all the time. This noise is spontaneous and random, and it does not really interfere with the estimation of signal characteristics. Some examples of sensory noise in the visual domain: bad lightning conditions, visual reflections; some examples of sensory noise in the auditory domain: sound reverberations, changes in sound travel due to temperature or humidity. On the other hand, *permanent conflicts* are also experienced, when there is a systematic bias in one or both of the modalities. A natural example is growth when the physical properties of the body changes. In the visual domain, it affects the length or the separation of the eyes; in the auditory domain it affects the inter-aural difference altering the binaural cues or the shape of the pinnae altering the monaural cues. A sensory handicap also results in a systematic, long-lasting bias.

Now, we are discussing the characteristics of the ventriloquist aftereffect. The magnitude of the aftereffect is considered to depend mainly on two factors: i) duration of the exposure period, and ii) AV discrepancy during exposure. It is generally 10-50% of the AV conflict size, but it shows large inter-subject variability (Chen & Vroomen, 2013). The largest

effect can be observed in the spatial locations used during exposure, but to a smaller extent, adaptation also generalizes to untrained locations (Bertelson et al., 2006). There is no clear consensus about the generalization of aftereffects across frequencies. Whilst early studies did not find a transfer across frequencies (Lewald, 2002; Recanzone, 1998), Frissen and colleagues demonstrated transfer in a large population across a wide range of frequencies (from 400 Hz to 6400 Hz; Frissen et al., 2003, 2005). Traditionally, the ventriloquist aftereffect is obtained after an exposure period of several minutes (Canon, 1970; Radeau & Bertelson, 1974, 1977, 1978; Recanzone, 1998). Although, more recent evidence suggests that recalibration can occur much faster (Frissen, de Gelder, & Vroomen, 2012), or even after a single AV exposure (Mendonça et al., 2015; Wozny & Shams, 2011b). Interestingly, the recalibration does not dissipate quickly and stays on for minutes (Frissen et al., 2012), especially using short interleaved exposure periods during post-test (Wozny & Shams, 2011a).

Reliability weighting has been well established as an optimal strategy for integrating multisensory signals with the goal of maximising the precision of the unified percept. Similarly, reliability weighting has been proposed for AV recalibration (Ghahramani, Wolpert, & Jordan, 1997; Witten & Knudsen, 2005). Although, the reliable cue is not always accurate and inaccurate unisensory signals can easily lead to an inaccurate unified percept (Ernst & Di Luca, 2011). Crucially, The Gordian knot of accuracy cannot be determined directly from the sensory estimates, thus without external feedback or prior knowledge the nervous system is unable to recognise and ascertain the level of accuracy (Ernst & Di Luca, 2011; Zaidel, Ma, & Angelaki, 2013). Meanwhile, other studies suggested visual-dominant adaptation as a theoretical model in the AV (Knudsen, 2002; Spence, 2009) or other multisensory domains (Rock & Victor, 1964). In the visuo-haptic domain, developmental

studies demonstrated that vision dominates touch for orientation, whilst touch dominates vision for sizes (Gori, Del Viva, Sandini, & Burr, 2008; Gori, Sandini, Martinoli, & Burr, 2010). The results were interpreted as the more accurate modality recalibrates the other one highlighting that accuracy is not a general feature of the senses, but it appears in task contexts. They propose that recalibration might be more important than integration before the age of 8, given the nature of growth and the maturation of senses (Burr, Binda, & Gori, 2011).

The currently prevailing theoretical framework of adaptation based on accuracies was further corroborated by a recent visuo-vestibular recalibration study (Zaidel, Turner, & Angelaki, 2011). Moreover, the authors claimed that visual-dominant adaptation is only a subcase of fixed-ratio adaptation and they provided experimental evidence for the account of the more general fixed-ratio adaptation (Zaidel et al., 2011). They noted that although the senses achieve internal consistency through this unsupervised adaption, without external feedback the brain cannot establish a veridical representation. They tested external, supervised recalibration in a follow-up study, where the both the effects of reliability and accuracy were manipulated (Zaidel et al., 2013). They showed that the less reliable and more inaccurate cue gets recalibrated. Strikingly, they found that when the more reliable sensory signal was inaccurate, the senses were yoked and calibrated together in the same direction. They concluded that unsupervised and supervised calibration work in parallel, where unsupervised adaptation can calibrate the senses independently, but the supervised adaptation calibrates only based on the multisensory percept (Zaidel et al., 2013).

Neural basis of audio-visual perception

For many decades, research in neuroscience and psychology has been focusing on single senses e.g. vision and audition. *Functional specialization* was a common theme referring to the specialized functions of different brain regions e.g. visual processing in the occipital lobe

(Gazzaniga, 2000). Multisensory integration was considered to take place in higher-order association areas after extensive unisensory processing of the sensory signals (Felleman & Van Essen, 1991). Intriguingly, recent evidence based on anatomical, neurophysiological and imaging studies suggests that multisensory integration occurs already in the primary sensory cortices (Driver & Noesselt, 2008; Ghazanfar & Schroeder, 2006). In this section, I will focus on the neural mechanisms of audio-visual interaction, in particular, how auditory spatial processing is influenced by visual signals. At first, I will introduce how audio-visual information converges in subcortical and cortical multisensory regions by feedforward mechanisms. Secondly, I will demonstrate that unisensory processing can be modulated by feedforward and feedback mechanisms from other cortical regions. I will also present the neural substrates of the ventriloquist effect and its causal inference based on neuroimaging evidence. Finally, I will shed light on the coordinate transformations performed by the posterior parietal cortex to enable a common framework for audio-visual spatial representation.

The *superior colliculus* (SC) is the model structure of multisensory convergence (Meredith & Stein, 1983, 1986; Stein & Arigbede, 1972). The three general principles of multisensory integration, such as the spatial rule, the temporal rule and the principle of inverse effectiveness were described based on long-term studies on this subcortical structure (Stein & Meredith, 1993). In addition to the SC, other subcortical structures involved in audio-visual perception are the *basal ganglia* and the *putamen* (Stein & Meredith, 1993; von Saldern & Noppeney, 2013).

Now, we are turning to the classical cortical sites of feedforward audio-visual convergence. Direct neurophysiological (Barraclough, Xiao, Baker, Oram, & Perrett, 2005; Bruce, Desimone, & Gross, 1981) and numerous imaging studies (Beauchamp, Lee, Argall, &

Martin, 2004; Calvert et al., 1999, 2000; Stevenson, Geoghegan, & James, 2007) showed audio-visual convergence in the *superior temporal sulcus* (STS). Bidirectional anatomical connections with auditory and visual cortices have been also described (Padberg, Seltzer, & Cusick, 2003). Another well-known convergence region is *the intraparietal sulcus* (IPS) (R. A. Andersen, Snyder, Bradley, & Xing, 1997; Bremmer et al., 2001; Cohen & Andersen, 2004; Rohe & Noppeney, 2015a, 2016; Sereno & Huang, 2014) including various sub-regions e.g. the lateral intraparietal area (LIP) extensively studied in macaques (Cohen, 2009; Cohen & Andersen, 2004). The *temporo-parietal area* (TPT) located at the temporo-parietal junction is another cortical region involved in the representation of multimodal space (Leinonen, Hyvärinen, & Sovijärvi, 1980). The *ventrolateral prefrontal cortex* (VLPFC) has been investigated mostly in the last decade as another cortical region for audio-visual convergence (Barbas et al., 2005; Fuster, Bodner, & Kroger, 2000; Lizabeth M Romanski, 2007; Sugihara, Diltz, Averbeck, & Romanski, 2006). The *premotor cortex* is another frontal region receiving inputs from auditory and visual regions (Bremmer et al., 2001; Graziano, Reiss, & Gross, 1999; Graziano, Yap, & Gross, 1994; Russo & Bruce, 1994). It is important to note that different methodologies have been used to assess audio-visual integration in the aforementioned studies from anatomical tracing to electrophysiology and neuroimaging. These approaches evaluate integration at the level of neurons (electrophysiology) or at a mixture of neuronal populations (neuroimaging), so careful considerations are needed for interpretation and generalization of integration.

Several anatomical studies provided evidence that a reverse information flow is also present in audio-visual processing: feedback connections from association areas project to the primary and secondary auditory (Barnes & Pandya, 1992; Hackett, Stepniewska, & Kaas, 1998; L M Romanski et al., 1999; Smiley et al., 2007) and visual (Falchier, Clavagnier,

Barone, & Kennedy, 2002) cortices. In parallel, early functional imaging studies reported activation in the auditory cortex during silent lip-reading (Calvert et al., 1997), and modulations in the auditory and visual cortex by bimodal speech stimuli (Calvert et al., 1999). Interestingly, initial reports on auditory activations in the visual cortex appeared already 30-40 years earlier (Bental, Dafny, & Feldman, 1968; Fishman & Michael, 1973; Lomo & Mollica, 1959; Morrell, 1972; Murata, Cramer, & Bach-y-Rita, 1965; Spinelli, 1968), but the results received lots of scepticism and they were attributed e.g. to non-specific effects or confounding factors. This time though, the findings attracted much attention and the topic of *early sensory integration* received a surge of interest in the following years. A vast body of research confirmed that audio-visual interactions occur at the early sensory cortices using neuroimaging (Kayser, Petkov, Augath, & Logothetis, 2007; Lehmann et al., 2006; Martuzzi et al., 2007; Miller, 2005; Pekkola et al., 2005; van Atteveldt et al., 2004; Watkins, Shams, Tanaka, Haynes, & Rees, 2006) and neurophysiological (Bernstein, Auer, & Takayanagi, 2004; Besle, Fort, Delpuech, & Giard, 2004; Bizley, Nodal, Bajo, Nelken, & King, 2007; Brosch, Selezneva, & Scheich, 2005; Fu et al., 2004; Ghazanfar, Maier, Hoffman, & Logothetis, 2005; Giard & Peronnet, 1999; Molholm et al., 2002; Schroeder & Foxe, 2002; Schwartz, Berthommier, & Savariaux, 2004; van Wassenhove, Grant, & Poeppel, 2005) methods. Schroeder and Fox investigated more directly the nature of the modulatory mechanisms, and described both feedforward and feedback mechanisms that are distinctive in their laminar profiles (Schroeder & Foxe, 2002). Meanwhile, anatomical studies revealed lateral cross-connections from the primary auditory cortex to the primary and secondary visual cortices in the macaque monkey (Falchier et al., 2002; Rockland & Ojima, 2003) and similar lateral connections were found from the visual cortex to the auditory cortex in the ferret (Bizley et al., 2007) and the Mongolian Gerbil (Budinger, Heil, Hess, & Scheich, 2006).

These results opened a whole new era in our understanding of multisensory processing, and in their seminal paper, Ghazanfar and Schroeder even proposed intriguingly that the neocortex is essentially multisensory (Ghazanfar & Schroeder, 2006). For recent developments in the topic I refer to some more recent reviews e.g. (Driver & Noesselt, 2008; Kayser, Petkov, Remedios, & Logothetis, 2012; Musacchia & Schroeder, 2009). Now, we are changing focus to the neural substrates of the ventriloquist effect.

The ventriloquist effect has been the most studied behavioural paradigm in the multisensory field; nevertheless, the possible neural mechanisms remained unknown for a long time. The first imaging studies described indirect findings related to the immediate ventriloquist effect (Bischoff et al., 2007; Colin, Radeau, Soquet, Dachy, & Deltenre, 2002; Gondan, Niederhaus, Rösler, & Röder, 2005; Stekelenburg, Vroomen, & de Gelder, 2004; Teder-Sälejärvi, Russo, McDonald, Hillyard, & Di Russo, 2005), but they did not find direct evidence for a visual influence on the auditory percept in the auditory cortex. The first direct evidence came from an EEG-fMRI study of Bonath and colleagues (Bonath et al., 2007) describing a bias in the left-right balance in auditory cortex activity during the illusion. The finding was extended in a follow-up study investigating the modulatory effects of asynchrony on the ventriloquist effect. The authors demonstrated that a change in left-right balance of the neural activity could be localized to the planum temporale (PT) (Bonath et al., 2014). Another study corroborated the results, and demonstrated that the spatial representation of unisensory auditory stimuli as well as the illusory percept can be measured by BOLD response changes, which supports the population code hypothesis (Callan, Callan, & Ando, 2015). These results are also in great agreement with the hemifield code hypothesis of auditory spatial representation (McAlpine, 2005; Ortiz-Rios et al., 2017; Salminen, May, Alku, & Tiitinen,

2009; Stecker, Harrington, & Middlebrooks, 2005) that has been corroborated in a very recent high-resolution fMRI study (Ortiz-Rios et al., 2017).

CHAPTER 2: METHODOLOGICAL FOUNDATIONS

In the current chapter we introduce the methods that form the basis for stimuli, tasks, designs and analysis approaches used frequently throughout the thesis. At first, we introduce the cues humans utilize for spatial hearing, in particular the role of head-related transfer function and how it can be used for auditory stimulus presentation. Subsequently, we discuss psychophysical procedures that are useful for evaluating sound localization abilities of observers and determining threshold performance. Finally, we delve into the details of multi-variate pattern analysis, a very powerful technique to characterize representations of neural activation patterns.

Head-related transfer function

The brain combines multiple cues to identify the sources of sound signals. The cues can be classified as binaural cues if their processing depends on both ears or monaural cues if they can be processed by one ear (Moore, 2013). Binaural cues are most useful in sound localization along the horizontal plane (left to right), whilst monaural cues are needed to help localization along the median plane (upside down and front to back) (R A Butler, 1969; Middlebrooks & Green, 1991; Wightman & Kistler, 1997).

The two most prominent *binaural cues* are the intensity and time differences sensed at the ear, called inter-aural level difference (ILD) and inter-aural time delay (ITD), respectively (Middlebrooks & Green, 1991). Low-frequency sounds have a wavelength similar to the size of the head, therefore they can “bend around” the head resulting in negligible ILD changes below 500 Hz; on the contrary, shadowing effects can be as large as 20 dB at higher frequencies (Feddersen, Sandel, Teas, & Jeffress, 1957). ITD can be calculated from the path difference between the two ears as illustrated by Figure 2.1. The maximum ITD is reached by

a sound source opposite to one ear, where the expression results in about 690 μs. This number corresponds well to experimental data (Feddersen et al., 1957). For a pure tone, ITD appears as a phase difference between the two ears, and as the maximum delay approaches 180°, the phase difference becomes ambiguous limiting ITD cues mainly below 1500 Hz. The theory that ITDs can be detected and utilized at low frequencies, whilst ILD cues are most useful at higher frequencies was first proposed by Lord Rayleigh (Rayleigh, 1907). Although, the theory was originally based on pure tones, but it holds reasonably well for complex sound signals (Moore, 2013).

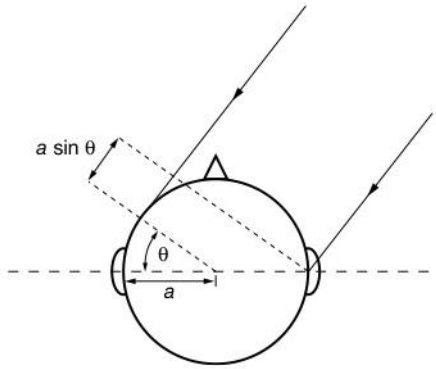


Figure 2.1 Method for calculating ITD from path difference. A distant sound source arriving to the two ears at an angle θ and assuming spherical head with radius a (9 cm) results in a path difference given by $d = a\theta + asin\theta$. Subsequently, the time delay is calculated by $ITD = dv$, where v denotes the speed of the sound (30 μs/cm).

Monaural cues are mainly based on the change in the spectral properties of the sound due to the pinna, the head and the torso. They act as a spectral filter reinforcing or attenuating the sound on particular frequencies depending on their shape and the direction of the incoming sound relative to the head. Spectral changes based on the pinna are limited to frequencies above 6000 Hz, since sound waves at these frequencies have sufficiently short

wavelengths to be able to interact with the pinna. However, spectral changes can occur also at much lower frequencies due to effects of the shape of the head and the torso (Blauert, 1969; Robert A Butler, 1971).

Spatial cues of sound sources can be technically characterized by measuring the spectrum of the sound at the ear drum relative to the original source. The ratio of the two is called the *head-related transfer function* (HRTF). Importantly, HRTFs incorporate all the spatial cues that are accessible to the observer, therefore they are the ideal and most compact way of characterizing spatial cues of sound sources. Measuring HRTFs of an individual is a laborious task, and it involves measurements of transfer function at many spatial directions around the head as well as removing artefacts due to the transfer functions of the setup (loudspeaker, microphone). There are two practical ways around this tedious work. One approach is to use pseudo-individualized or standard HRTFs instead of individualized ones. The other approach is to use binaural recordings similarly as done with HRTFs, but only at specific sound locations and without engineering the sounds to remove artefactual transfer functions due to the setup.

The individual differences of HRTFs and their importance in sound localization has been the subject of multiple studies (Kawaura, Suzuki, Asano, & Sone, 1991; Middlebrooks, 1999; Moller, Sorensen, Jensen, & Hammershoi, 1996; Wenzel, Arruda, Kistler, & Wightman, 1993). It has been demonstrated that humans made less errors via their own HRTFs than when they listened the sound sources via other's HRTFs (Middlebrooks, 1999; Moller et al., 1996). Yet, Wenzel and colleagues (Wenzel et al., 1993) and Kawaura and colleagues (Kawaura et al., 1991) showed earlier that horizontal localization is rather robust and non-individualized HRTFs provide sufficient cues for localization in the horizontal plane. There were efforts to make a standard HRTF that represents a typical subject in the population. Gardner and Martin

made such measurements on a mannequin, called KEMAR (Knowles Electronics Mannequin for Acoustic Research) (Gardner & Martin, 1995). The measurements were made both on a small pinna that is representative of the pinna dimensions in the population and a large pinna that can be used for extreme dimensions.

We used binaural recordings as well as standard HRTFs in our experiments. The binaural recordings were made in an anechoic chamber for each individual. The details of the recording process are described in the methods sections of the chapters, where recordings were used (Chapter 3, Chapter 4). The standard HRTFs with small pinna (Gardner & Martin, 1995) were used in the experiment of Chapter 5. Here, we provide pilot sound localization results with both approaches to provide evidence that they are appropriate for our behavioural and fMRI experiments.

For the evaluation of binaural recordings, six participants took part of short sessions in laboratory and scanner. Participants were presented with recorded auditory stimuli from ±10°, ±7°, ±5°, ±4°, ±3°, ±2°, ±1°, 0° visual angle in a forced choice left-right discrimination task. A cumulative Gaussian was fitted to the percentage 'perceived right responses' as a function of stimulus location (www.palamedestoolbox.org). For details about the task and psychometric functions, see the next section of *Psychophysical procedures*. Figure 2.2 A-B show the group level psychometric functions. Paired t-test on the slope and threshold estimates did not reveal any significant difference supporting evidence for the usage of binaural recordings in scanner environment similarly as in laboratory.

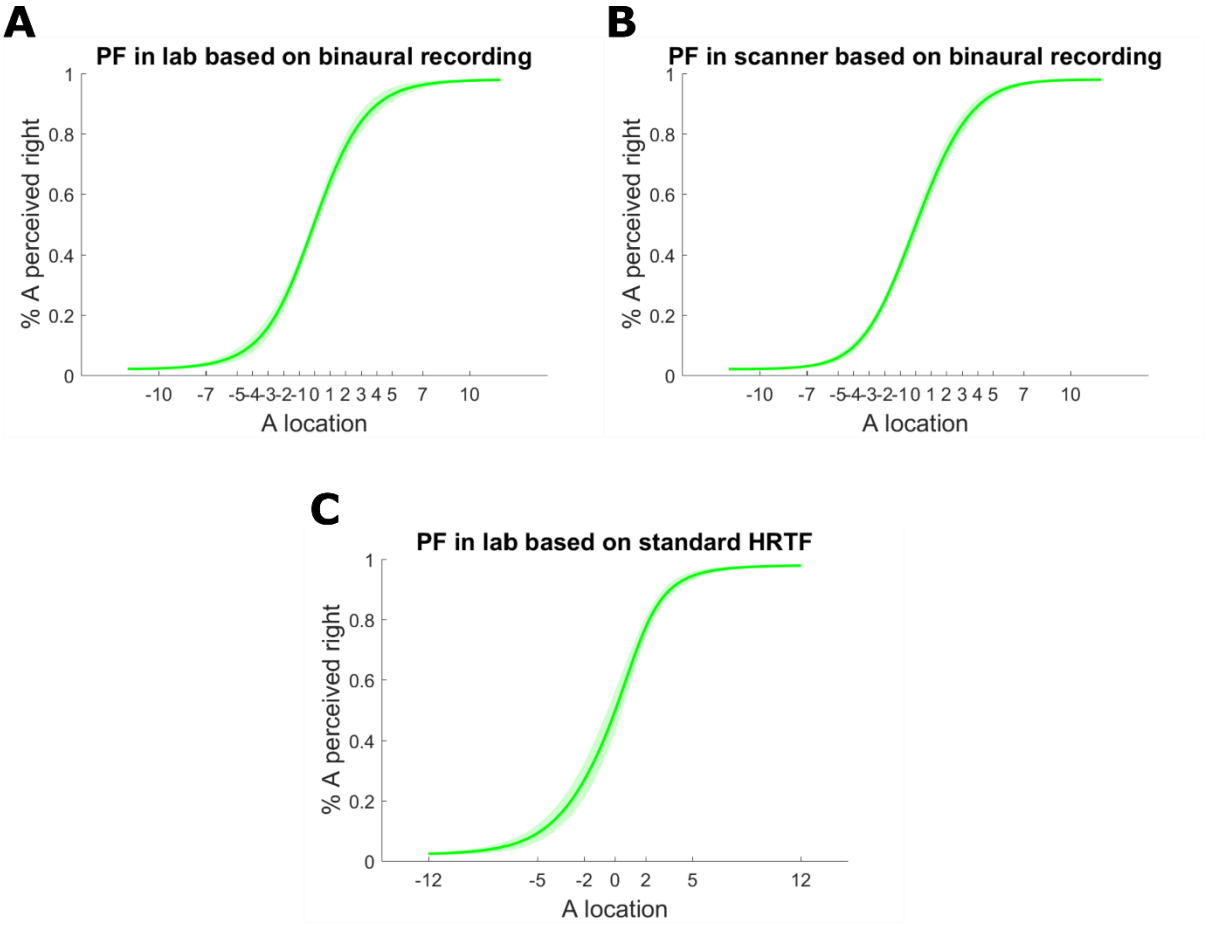


Figure 2.2 Pilot psychometric functions in laboratory and scanner using binaural recordings and standard HRTFs. (A-B) Psychometric functions in laboratory (A) and scanner (B) based on the mean fits of six individuals using binaural recordings. **(C)** Psychometric functions in laboratory based on mean fits of ten individuals using standard HRTFs. Shaded areas represent ±SEM in all panels.

For the evaluation of standard HRTFs, ten participants took part in a sound localization task as part of a pilot session for the behavioural experiment in Chapter 5. Participants were presented with recorded auditory stimuli from ±12°, ±5°, ±2°, 0° visual angle in a forced choice left-right discrimination task, and psychometric functions were fitted as before. Figure 2.2 C shows the group level PF. Two sample t-test on the slope and

threshold estimates between the binaural recordings (both in laboratory and scanner) and standard HRTFs did not reveal any significant difference suggesting that they are similar approaches in our experimental settings.

Psychophysical procedures

The beginning of psychophysics dates back to the 19th century, when Gustav Theodor Fechner set out the principles in his book, *Elements of Psychophysics* (Fechner, 1860). Psychophysics aims to quantify the relationship between physical stimuli and their perceptual counterparts (for a practical introduction, see Kingdom & Prins, 2010). There are various tasks and methods to characterize this relationship. Performance-based tasks measure 'how well' an observer performs in a particular task. On the other side, appearance-based tasks measure the stimulus-perception relationship without an explicit judgement of the performance and focuses on the stimulus appearance. This section deals with performance based tasks using threshold based approaches. In particular, we discuss forced-choice procedures, where observers' make a choice between two pre-specified options in a discrimination task.

Psychometric function

Psychometric function (PF) is a model with a sigmoidal shape describing the relationship between a stimulus level (e.g. spatial auditory location) and the performance in a forced-choice task (e.g. rightward responses in a left-right discrimination task) (Wichmann & Hill, 2001a, 2001b). Critically, one determines parameters of the PF to summarize behaviour, hence, the generic formulation of the psychometric function:

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

where γ denotes the guess rate (probability of correct response/discrimination when the stimulus is not detected), λ denotes the lapse rate (probability of incorrect

response/discrimination independent of stimulus level), α denotes the threshold (point of subjective equality in a discrimination task), β denotes the slope (rate of change of the function). Guess and lapse rates are generally not of interest, therefore they are kept as fix parameters during estimation (a small, non-zero value to allow some flexibility for PF fitting resulting in more reliable estimates of α and β) (Kingdom & Prins, 2010). Interestingly, we note that there is a third possibility in parameter estimation, namely constraining a parameter in multi-condition fitting. In this case the parameter (e.g. slope) will be estimated in all conditions at the same time, whilst its value is kept constant across conditions.

Five functions are commonly used to model psychometric data: Cumulative Normal, Logistic, Weibull, Gumbel and Hyperbolic Secant function. Probably, the Cumulative Normal Distribution is the theoretically most justified model that can be derived from the central limit theorem (Hays, 1994) assuming that a linear combination of independent noise sources underlie the decision process. The Cumulative Normal distribution function is formulated as:

$$F(x;\alpha,\beta) = \frac{\beta}{\sqrt{2\pi}} \int_{-\infty}^{x} exp\left(-\frac{\beta^2(x-\alpha)^2}{2}\right)$$

Interestingly, in this case the inverse of the slope parameter (1/β) is equivalent to the standard deviation of the PF that is also equivalent to the just noticeable difference (JND) defined at 84% level. Figure 2.2 illustrates a group level PF fitted by a Cumulative Gaussian function.

In our experiments, PFs are fitted to stimuli presented using the method of constants. In this popular method, the stimulus levels are randomized before the experiment and presented in the predefined order during the experiment. The obvious advantage of the approach is its simplicity and capability to estimate both threshold and slope parameters. On the other hand, pilot work is needed to determine the right stimulus levels, otherwise presentation of irrelevant stimulus levels might be time consuming and it can lead to poor

parameter estimates. An ideal range for the stimulus levels should be chosen such that the PF goes from just above chance (guess rate) to almost 100% correct/discrimination and only 1-1 stimulus level should produce these values. Typical psychophysics experiments contain 4 to 10 stimulus levels with 20 to 100 trials at each level (Wichmann & Hill, 2001b). The more the trials, the better the estimation process, however, sometimes one is interested only in the threshold parameter, when less stimulus repetitions might also yield to a sufficient estimate.

We also note that PFs can be fitted to stimuli presented during an adaptive procedure, either only for threshold (e.g. Quest by Watson & Pelli, 1983) or threshold and slope estimation at the same time (e.g. Psi method by Kontsevich & Tyler, 1999). In both cases, PFs are fitted iteratively after every trial and the produced estimates are used for presenting the next stimulus.

Generally, there are two procedures for fitting psychometric functions. The simpler choice is maximum likelihood (ML) estimation. During ML estimation parameters are searched in a way to maximize the likelihood of generating the data. Another procedure for fitting is Bayesian estimation that combines ML estimation with prior distributions of the parameters yielding posterior estimates using Bayes rule.

Adaptive staircase method

The purpose of adaptive staircase methods is to increase efficiency of the testing procedure. It is achieved by updating the presented stimulus levels based on observer's previous responses and converging to a threshold in a staircase procedure. Staircase procedures were developed originally by Dixon and Mood (Dixon & Mood, 1948)and use an up/down rule. Namely, a staircase procedure decreases subsequent stimulus levels after incorrect and increases subsequent stimulus levels after correct responses of the observer. The original up/down method is based on the last trial and uses the same step sizes for level the stimulus up or

down, therefore targeting thresholds at 50% correct responses. More sophisticated up/down methods were proposed later taking into account more preceding trials (transformed up/down method, (Wetherill & Levitt, 1965)) or using different sizes for step up and down (weighted up/down method, (Kaernbach, 1991)). Of course, the combination of the latter two methods can be also used and was proposed by García-Pérez (García-Pérez, 1998) in the transformed and weighted up/down method. The targeted proportion correct response can be calculated as:

$$\psi_{target} = \left(\frac{\Delta^+}{\Delta^+ + \Delta^-}\right)^{\frac{1}{D}}$$

where ψ_{target} is the targeted proportion correct, $\Delta^+$and Δ^- are the step up and down sizes, respectively, and D is the number of consecutive responses after which a step down is made.

Several methods exist to evaluate the convergence of staircase procedures. The most common approach is to terminate the staircase after a specific number of revearsals of direction have occurred (García-Pérez, 1998). In this case, the threshold is calculated as the average stimulus across the last trials a reversal occurred. Another option is to terminate after a specific amount of trials have occurred and the threshold is calculated on a specified amount of last trials. The second approach might not yield in a robust convergence, however, it has the benefit of a fixed amount of trials, therefore might be a faster approach. Typically, one uses multiple staircases runs and the final threshold estimate is an average of the run-specific estimates. Finally, we should mention that a hybrid approach also exists, where staircase runs are used to select stimulus intensities, and a PF is fitted to the data yielding the threshold estimate (Hall, 1981). Obviously, as we demonstrated in the previous section (Method of constants using psychometric function fitting), this strategy has the obvious disadvantage of assumptions to be made about the specific function and possibly, some of its parameters.

Signal detection theory

A qualitatively different approach to measure observers' performance in a psychophysics experiment is based on Signal Detection Theory (SDT). This method enables one to evaluate performance taking into account response biases. The main assumption behind the theory is that observers take decisions on the basis of information coming from two distributions: a signal distribution and a noise distribution. The observer has to make a forced choice between these distributions. A 'yes-response' to a signal coming from the signal distribution is a hit, whilst a 'no-response' from the same signal distribution is a miss. On the other hand, a 'yes-response' to a signal coming from the noise distribution is a false alarm, whereas 'no-response' to the same signal is a correct rejection. See Figure 2.3 for an illustration.

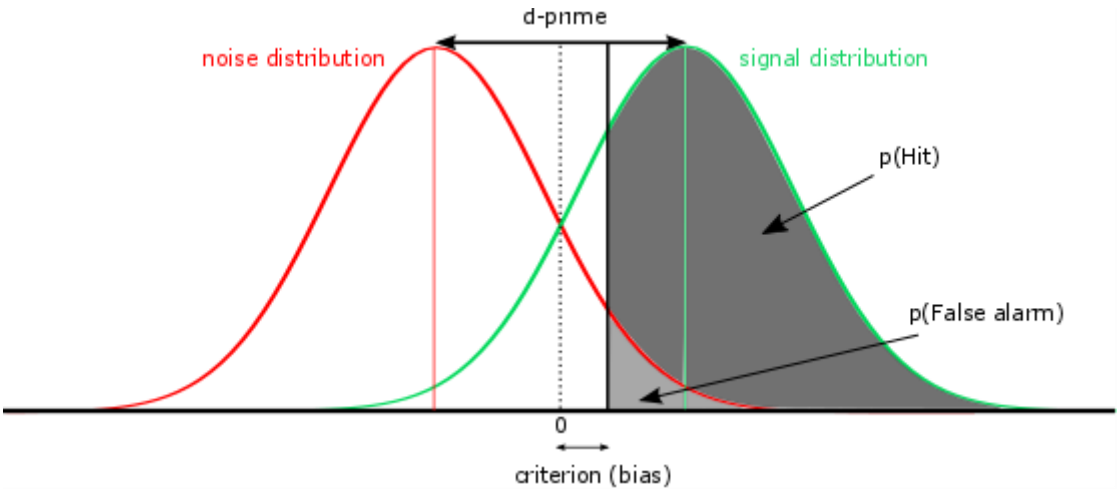


Figure 2.3 Signal and noise distributions with measures of d', criterion, hit rate and false alarm rate.

The measure which combines the likelihood of hits with the likelihood of false alarms is the observer's sensitivity, or d'. The d' measures how well the observer can tell apart signals coming from the signal and the noise distributions. The formal calculation of the d' is given by:

$$d' = z(Hit\ rate) - z(False\ alarm\ rate)$$

where the function z(x) is the inverse of the Cumulative Gaussian function. Larger values of d' indicate that the observer is better in discriminating signals coming from the signal and the noise distributions. A second SDT measure is the criterion. This measure represents the observer's response bias. The formal calculation of the criterion is given by:

$$criterion = -\frac{\big(z(Hit\ rate) + z(False\ alarm\ rate)\big)}{2}$$

A criterion = 0 means that the observer is unbiased, whereas a criterion different from 0 indicates that the observer is biased towards one of the two distributions (for an illustration, see below).

Multivariate pattern analysis of fMRI data

Conventional fMRI analysis focuses on brain regions that are involved in specific cognitive tasks (Friston, Holmes, Poline, et al., 1995; Friston, Holmes, Worsley, Frith, & Frackowiak, 1995; Friston, Jezzard, & Turner, 1994; Worsley & Friston, 1995). In order to characterize activation in the involved brain regions, data is normally spatially smoothed and the activation is averaged within a region of interest or cluster. In recent years, a growing number of studies go beyond this macroscopic characterization and target the information content represented in activation patterns of brain regions or the whole brain (Allefeld & Haynes, 2014; Cox & Savoy, 2003; Haxby, 2001; Haynes & Rees, 2006; Kriegeskorte, Goebel, & Bandettini, 2006; Norman, Polyn, Detre, & Haxby, 2006; Pereira, Mitchell, & Botvinick, 2009; Tong & Pratte, 2012; Walther et al., 2016)

There are three qualitatively different question that one might ask about activation patterns: (i) 'is there any information of interest' (pattern discrimination); (ii) 'where is the information' (pattern localization); (iii) 'how is the information encoded' (pattern

characterization) (Pereira et al., 2009). Most of the multi-variate pattern analysis (MVPA) studies target the first two questions: they aim to discriminate between stimuli or other experimental conditions and possibly, also to localize these effects. At first, we present the classification problem, and we outline a typical MVPA decoding analysis. Secondly, we describe the various steps in details, we show how these two questions can be tackled during the decoding process, and discuss the various options at each processing stage. With regards to the third question, we refer to excellent reviews on encoding and representational models from recent years (Diedrichsen & Kriegeskorte, 2017; Mur, Bandettini, & Kriegeskorte, 2009; Naselaris, Kay, Nishimoto, & Gallant, 2011).

The essence of MVPA decoding is the classifier: a function that finds a mapping between patterns of features to given labels to separate examples belonging to different conditions. Labels can be discrete classes (e.g. stimulus left and right) or continuous variables (e.g. stimuli at various visual angles). In the previous case the MVPA formulation is called a classification problem, in the latter case it is called a regression problem. The data representation is illustrated in Figure 2.4.

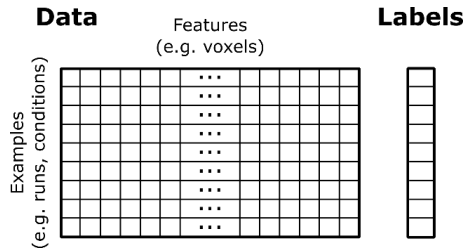


Figure 2.4 Data representation for MVPA analysis. Each row represents an example with features as voxels and a discrete or continuous label. Different rows represent examples belonging to different conditions, or multiple instances (e.g. different runs) of the same condition.

A key step in the decoding analysis is partitioning the dataset into independent training and test sets. Once the classifier is trained to find a mapping in the training set, it is applied to the test set to predict labels for the test examples. Then the true labels are compared to the predicted labels and the accuracy (for classification) or mean squared error/correlation (for regression) is calculated as a measure for decoding. For testing against chance-level decoding, permutations tests have been proven to be the valid statistical approach in the last years (Allefeld, Görgen, & Haynes, 2016; Pereira & Botvinick, 2011; Stelzer, Chen, & Turner, 2013). Now, we discuss the steps of the decoding process in details:

1. Creating examples: the first important choice is how to identify examples. Examples can be created commonly from raw fMRI volumes, single trials, blocks or runs. Some level of temporal aggregation is almost always needed to increase the signal-to-noise level and the particular choice depends on the specific design. Blocked designs or sparse event-related designs are suitable for examples to be created from raw data, whilst a classical GLM estimation is needed for rapid event-related designs due to the overlapping BOLD responses of subsequent trials. In the latter case, a common choice is to define a regressor for each condition in a run, although, we note that in case of conditions with many trials, it might be beneficiary to define multiple regressors per condition. In both cases, the regressors will be turned into examples after GLM estimation. One important detail between working from raw fMRI data and GLM estimates is that in the latter case one might specify additional nuisance regressors (e.g. for movement) that explain some of the variability in the data resulting in better estimates of interest.
2. Feature selection: the obvious choice for fMRI data that features are defined as voxels in a vectorised format, however, it is less obvious at what scale they should be defined. Whole brain analyses suffer from large number of noisy voxels, therefore dimensionality

reduction techniques are generally used in this case. Common approaches to reduce the number of features are to perform PCA (Hansen et al., 1999), ICA (Calhoun & Adali, 2006) or recursive feature elimination (Hanson & Halchenko, 2008). Most research questions are interested in localization of patterns, therefore a more suitable approach is to reduce the voxels to region of interests (ROIs) or to perform searchlight analysis. Importantly, these localized approaches reduce the concern that the classifier finds patterns by combining information from functionally distinct brain regions, since reflecting operations that are not actually performed by the brain. ROI based decoding is a very powerful approach, the only concern is the need of ROIs based on prior knowledge. When prior information is not available, searchlight analysis is a good choice, where decoding is iteratively performed on local clusters (typically spheres with radius of e.g. 15 mm) and sampled throughout the whole brain (Kriegeskorte et al., 2006). A disadvantage of the searchlight approach is the need to correct for multiple comparison, therefore they are commonly used together with group-level statistical inferences.

3. Example/feature normalization: feature normalization is a standard practice in the machine learning community. In addition, some authors in the neuroimaging community proposed that example normalization might be also beneficial for fMRI data (Pereira et al., 2009). Indeed, Euclidean normalization of examples is a default approach in the Pronto toolbox (Schrouff et al., 2013), and our results proved the beneficence of the approach. Very recently, it has been proposed to apply multivariate feature normalization based on the residuals from GLM estimation instead of the standard univariate feature normalization (Walther et al., 2016). The theoretical justification for the approach is that it leads to the pre-whitening of examples by removing spatial correlations left in the

residuals. The new technique appears to be welcomed in the neuroimaging community and it is already offered as a pre-processing step in the latest version of the popular The Decoding Toolbox (TDT) (Hebart, Görgen, Haynes, & Dubois, 2015).

4. Choice of classifier: various classifiers are at disposal for the decoding analysis. The simplest algorithm is called nearest neighbour that does not even learn a mapping function, but classifies based on finding the most similar training example (neighbour). It tends to work well only with a small number of features and generally applied with feature elimination methods (Haxby, 2001; Mitchell et al., 2004). Popular choices for more complex algorithms that do learn a mapping function are linear Support Vector Machines (SVM), Fisher's Linear Discriminant Analysis (LDA). and Gaussian Naïve Bayes (GNB) (Pereira et al., 2009). GNB is a good candidate for searchlight analysis (Pereira & Botvinick, 2011), although it is inferior to SVM in case of large number of features due to the regularization term of SVM helping to weigh down the effects of noisy features. Classifiers with a mapping function can be divided to generative and discriminative models (Hastie, Friedman, & Tibshirani, 2001). LDA is a generative approach assuming multivariate Gaussian distribution of the classes with separate means, but same within-class covariance matrix. Consequently, LDA performs well when the assumptions are hold and the covariance matrix can be estimated reliably (typically, in case of small number of features). On the contrary, linear SVM is a discriminative approach using regularization and with no distributional assumptions. It is a large margin classifier finding a subset of examples, called support vectors defining the class-separating hyperplane. Linear SVM is a very popular choice of algorithm in numerous MVPA toolboxes in recent years, see e.g. Pronto (Schrouff et al., 2013), TDT (Hebart et al., 2015) and PyMVPA (Hanke et al., 2009).

5. Training and testing using cross-validation (CV): CV is a standard approach in decoding to partition data into training and test sets. The most common variant is leave-one-run-out (LORO) CV. It splits the dataset for training and test sets as many times as runs are available leaving always a different run out for testing and all the others for training the algorithm. The final decoding measure (e.g. accuracy) is calculated as an average across CV steps. The advantage of LORO-CV for partitioning the data is two-fold: (i) the classifier can be trained with most of the data; (ii) the trained models can be tested on all data (of course, separately in different folds). The only disadvantage of LORO-CV is that it is computationally expensive. A compromise is to use k-fold CV, splitting the data to k folds (e.g. k=5 or10) and repeating the procedure similarly as for LORO-CV. It is essential in k-fold CV, that examples that are correlated (e.g. due to auto-correlation in the same run) are always kept in the same fold. Otherwise the classifier might predict examples only based on the existing correlation resulting in an false overestimation of the real decoding measure (e.g. accuracy) (Pereira et al., 2009). Figure 2.5 illustrates a 5-fold CV.

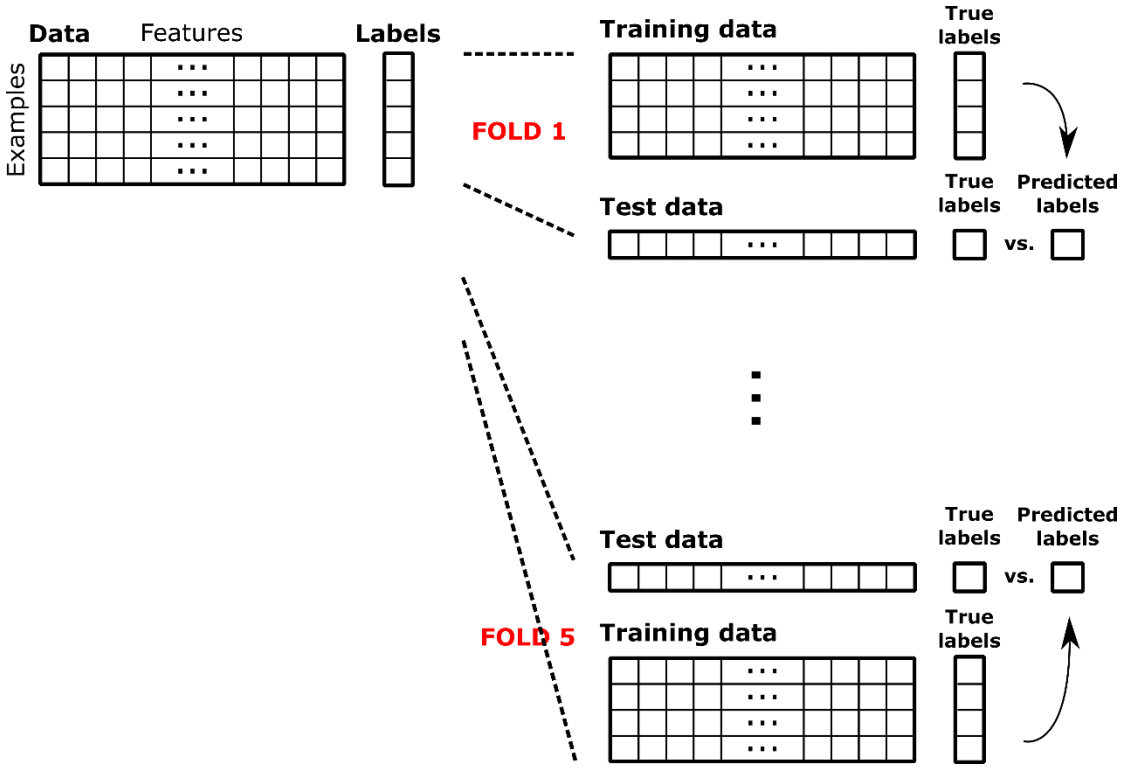


Figure 2.5 CV scheme for MVPA analysis. In this 5-fold CV, data are split into 5 folds. In each step, the classifier is tested on one left-out run, and trained on all the other runs. The final decoding measure (e.g. accuracy) is an average value across folds.

6. Statistical inference: similarly to classical inference of fMRI data, inference of decoding results can be done at two levels. At the subject level, true decoding measures can be compared to chance-level values. In recent years, permutation-tests have proven to be the valid statistical approach over t-tests or binomial-tests regarding information-like measures (Allefeld et al., 2016; Haynes, 2015; Nichols & Holmes, 2002; Pereira & Botvinick, 2011; Stelzer et al., 2013). Subject-level permutations are achieved with permuting labels within a given chunk of data (keeping correlated examples together is essential also here) and obtaining a distribution of permuted decoding measures. The true decoding values then can be compared to the permutation distribution. Permutation testing has only weak distributional assumptions and it is useful to reveal biases in the

decoding process (e.g. due to violation of independence between training and test set) that can lead to above-chance level baseline decoding. Permutation tests can be also applied to make inference at the group level (Allefeld et al., 2016; Stelzer et al., 2013). In this case, subject level permutations are bootstrapped or permuted at the second level calculating second level permutation distributions.

CHAPTER 3: NEURAL BASIS OF EXPLICIT CAUSAL INFERENCE IN AUDIO-VISUAL PERCEPTION

Introduction

Events in the environment most often generate signals in multiple senses at the same time. In order to make a veridical representation of the world, the brain needs to integrate sensory signals generated by the same event and segregate those generated by different events (Trommershäuser et al., 2011). Extensive research in the last decades has shown that there are three main bottom-up inter-sensory cues for such a decision: spatial disparity, temporal synchrony and semantic relations (Chen & Vroomen, 2013; Laurienti et al., 2004; HweeLing Lee & Noppeney, 2011; Recanzone, 2009; Slutsky & Recanzone, 2001; Welch, 1999). Accumulating evidence suggests that human observers arbitrate between sensory integration and segregation in line with Bayesian causal inference as illustrated also by spatial ventriloquism (Körding et al., 2007; Shams & Beierholm, 2010). In spatial ventriloquist paradigms, observers are presented with synchronous, yet spatially disparate audio-visual (AV) signals and report the perceived location of the stimuli and/or their judgement about a common or a separate underlying sensory source. The perceived location of the AV (or A in sound localization task) signal is based on the weighted reliability of the unisensory signals (Alais & Burr, 2004; Ernst & Banks, 2002), resulting in the mislocalization of the A signal in most of the natural conditions due to the superior reliability of the V signal (Howard & Templeton, 1966).

Recent neuroimaging research suggests that Bayesian causal inference emerges at different levels of the human cortical hierarchy (Rohe & Noppeney, 2015a). This work showed that, firstly, low-level sensory regions encoded primarily the location of the

corresponding A and V signals. Secondly, spatial estimates of AV signals were weighted based on their relative reliability in the posterior part of the intraparietal sulcus (IPS0-2). Critically, at the top of the hierarchy in the more anterior parts of IPS (IPS3-4), spatial estimates were formed by taking into account the uncertainty of the causal structure of the signals. Moreover, the authors demonstrated that IPS3-4 integrated AV signals weighted by their bottom-up reliabilities and top-down task-relevance (Rohe & Noppeney, 2016).These studies demonstrate that IPS3-4 forms spatial representations of AV sensory signals relying implicitly on causal inference (Rohe & Noppeney, 2015a, 2016). One question remains: which brain regions perform explicit causal inference Is it IPS3-4 performing implicit and explicit causal inference or are there other cortical regions that are able to estimate the probability of signals representing the same event and then pass the information to IPS3-4?

To address this question, human observers were presented with synchronous auditory and visual signals that were either spatially collocated (congruent) or discrepant (incongruent). On each trial, participants determined whether the AV signals were generated by same or separate events. We adjusted the AV disparity individually for each participant, such that they were approximately 70% correct in their causal judgment both for congruent and incongruent trials. Critically, this individual adjustment allowed us to dissociate between physical and perceived AV congruency. Based on previous studies demonstrating that arbitration between audio-visual integration and segregation can be related to the prefrontal cortex (Gau & Noppeney, 2016; Noppeney, Ostwald, & Werner, 2010), and given the hierarchical organization of the human brain, we expected that the dorsolateral prefrontal cortex (DLPFC) might play a key role in explicit causal inference, determining whether A and V sensory signals come from common or separate events. This decision would then modulate

in a top-down fashion the spatial representations of AV signals in other cortical regions (e.g. IPS or even lower level auditory and visual regions).

Methods

Participants

Twelve right-handed participants (11 females, mean age: 21.0; SD=2.9) gave informed consent to take part in the fMRI experiment. Two participants were excluded because their visual regions could not be reliably defined based on the retinotopic localizer scans acquired after the main experiment. One additional participant took part only in the retinotopic localizer session and did not progress to participating in the main experiment.

All participants were selected prior to the main experiment 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) unbiased sound localization performance outside and inside the scanner; and (v) ~70% accuracy during an initial screening procedure for the main task at an individually adjusted audio-visual disparity in a prior screening session outside the scanner. The study was approved by the human research ethics committee at the University of Birmingham.

Experimental procedure

Typically, participants completed 7 sessions. The most important details of each session are illustrated in Figure 3.1. The first session was completed in an anechoic chamber and lasted for ~1 hour. It consisted of the recording of sound stimuli used later in the experiment, and the initial assessment of participants' sound localization performance. Participants were trained in two other sessions to determine subject-specific AV spatial disparities in a mock scanner. The two sessions lasted for an overall ~2 hours. In the fourth session participants were moved to the scanner, where the adjusted AV disparities and the sound localization performance were

finalized. 1-2 more training sessions were provided for 3 participants, who were not able to start the main fMRI experiment in less than 3 weeks after the first training session. After the training, participants performed 2 sessions of the main experiment in the scanner, each lasting for ~1.5 hours. To enable the retinotopic mapping of visual and parietal cortical areas, participants performed a separate fMRI session with a standard retinotopic localizer task lasting for ~1 hour. 6 participants performed the retinotopic session before the main fMRI experiment, 4 participants performed the retinotopic session after the main fMRI experiment. 4 participants performed an additional behavioural session in the mock scanner after the main fMRI experiment to provide eyetracking data. Eyetracking recording was part of the two training sessions with the other 6 participants.

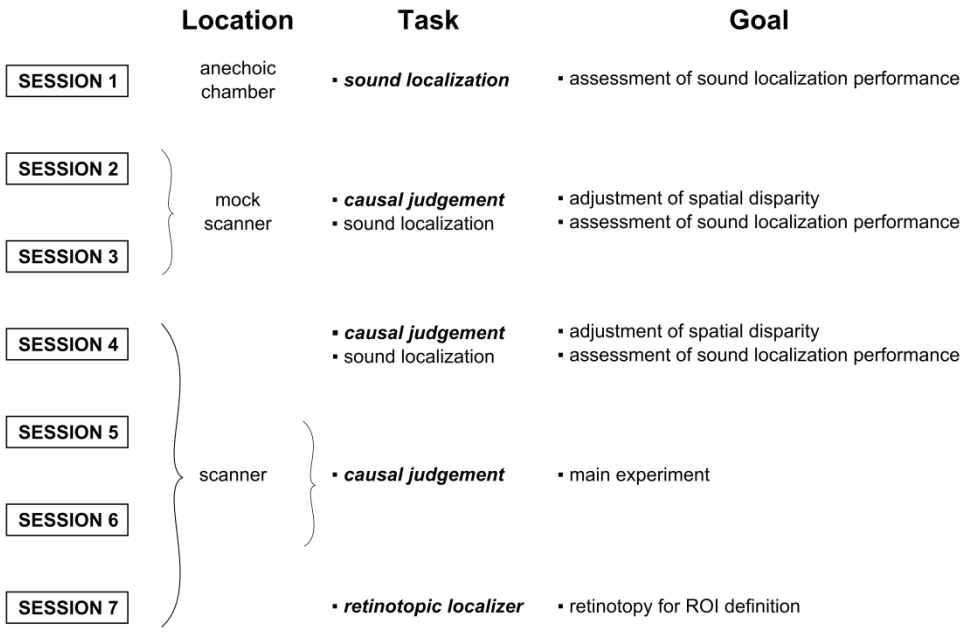


Figure 3.1 Typical experimental structure with 7 sessions. Experimental structure containing information about the most important details of the sessions: various tasks performed by the participants (major tasks of each session in bold) and the goal of the tasks in the experimental pipeline.

Stimuli

The visual stimuli were clouds of 20 white dots (diameter: 0.4° visual angle) sampled from a bivariate Gaussian presented on a dark grey background (70% contrast). The horizontal standard deviation of the Gaussian was set to a 5° visual angle, and the vertical standard deviation was set to a 2° visual angle. The sound stimuli were bursts of white noise with 5 ms on/off ramp. They were recorded individually for each subject with Sound Professionals™, Inc. (USA) in-ear binaural microphones in an anechoic chamber in the School of Psychology, University of Birmingham. The process consisted of displaying the sounds with an Apple Pro Speaker (at a distance of 68 cm from the participants) from -8° to 8° visual angle with 0.5° visual angle spacing, and at ±9° and ±12° visual angle along the azimuth. The participant's head was placed on a chin rest with forehead support and controlled by the experimenter to ensure stable positioning during the recording process. Five stimuli were recorded at each location (recording set) to ensure that sound locations could not be determined based on irrelevant acoustic cues. Importantly, visual stimuli were randomly generated on each trial, and auditory stimuli were randomly chosen from the recording set of five stimuli. This randomization step enabled that the decoding algorithm used for MVPA analysis will not get biased by specific stimulus characteristics, but will directly utilize the spatial stimulus features allowing generalization of the decoding results.

Assessment of sound localization performance – outside the scanner

In the first session , sound stimuli were recorded from a series of locations in an anechoic chamber as discussed in the stimuli section. Participants were presented with the recorded auditory stimuli from ±12°, ±9°, ±7°, ±5°, ±3°, ±2°, ±1°, 0° visual angle in a forced choice left-right discrimination task. A cumulative Gaussian was fitted to the percentage 'perceived right responses' as a function of stimulus location using maximum-likelihood estimation

(www.palamedestoolbox.org). The guess rate and lapse rate parameters (0 and 0.01, respectively) were kept fixed and the fitting procedure resulted in estimates of the threshold (point of subjective equality, PSE) and the slope (inverse of the standard deviation, STD) of the psychometric function. First, participants were familiarized with the recording process and the left-right discrimination task. Then the recording process and the left-right discrimination task were repeated multiple times until unbiased sound localization was performed on at least one recording set. Left-right discrimination tasks on 2 stimulus locations (sound stimuli at approx. at ±JND values) were also performed to help quick, initial evaluation of the recording sets. In the first session, unbiased sound localization was specified by PSE/STD ratios < 0.3.

In the following training sessions, sound localization was further assessed based on left-right discrimination task on 2 stimulus locations. Typically, 20-60 repetitions per stimulus location were performed in the mock scanner.

Adjustment of spatial disparity for each participant – outside the scanner

In order to dissociate spatial disparity (i.e. physical congruency) and perceived common source (or perceptual congruency), we employed a threshold approach: In two sessions we adjusted AV spatial disparity inside the mock scanner individually for each subject to obtain an accuracy of 70% on the main causal judgment task (i.e. common vs. separate sources). This individual adjustment of AV spatial disparity allowed us to compare physically identical AV signals that were perceived as coming from common or separate sources. In the initial session inside the mock scanner, subject-specific AV spatial disparities were adjusted in maximally 5 adaptive staircases, each targeting 70% accuracy on the common source judgment task (www.palamedestoolbox.org). The adaptive staircases were terminated after a minimum of 30 trials, when 8 reversals occurred within a window of 20 trials. The spatial disparity threshold averaged across the adaptive staircases formed the starting estimate for

additional manual fine tuning in subsequent runs of 60 trials where the AV disparity was held constant within a run and adjusted across runs in step size of 1-2° visual angles across runs to obtain a stable performance accuracy of 70%. Performance accuracy of 60-80% for one specific AV disparity (in a range from 4° and 16° visual angle) were used as a participant selection criterion to allow for comparison of physically identical trials AV that were perceived as perceptually congruent or incongruent.

A short second session was also held in the mock scanner on a separate day aimed to measure inter-session variability in the performance of the participants. During the second session, further fine tuning of AV disparities in subsequent runs of 60 trials was applied as before.

Final assessment of spatial disparity and sound localization – inside the scanner

To account for differences between the mock scanner and the real fMRI scanner and ensure spatially unbiased positioning of the participant inside the scanner, the AV spatial disparity was finally adjusted in additional 1-3 runs with constant disparity inside the scanner prior to the main common source judgment experiment.

Similarly, we assessed participants' sound localization in a left-right discrimination task on close to the 2 finally selected stimulus locations inside the scanner. Typically, 40-80 repetitions per stimulus location were performed in the scanner for each subject. Unbiased sound localization was defined as less than 30% difference in the accuracy for left and right side stimuli. Each participant of the main fMRI study completed at least 20 repetitions per stimulus location on the final auditory stimulus locations resulting in a group mean localization accuracy of 87% (SEM=0.02).

Main experimental design

In the main experiment, participants were presented with synchronous audio-visual (AV) stimuli of 50 ms duration at a stimulus onset asynchrony (SOA) of 2.3 sec. The auditory (A) and visual (V) stimuli were independently sampled from two symmetric spatial locations along the azimuth. On each trial, participants reported whether A and V signals were generated by common or separate sources. In addition, participants' hand responses were counter-balanced within subjects. Hence, the experimental design factorially manipulated: (i) visual stimulus location (left vs. right) (ii) auditory stimulus location (left vs. right), and (iii) motor response (left vs. right hand). For further analysis and characterization of the functional properties of the regions we categorized trials into: (i) physical congruency (i.e. AV congruent vs. incongruent); and (ii) perceived congruency (i.e. participant perceived and reported AV signals on a particular trial as common or separate sources).

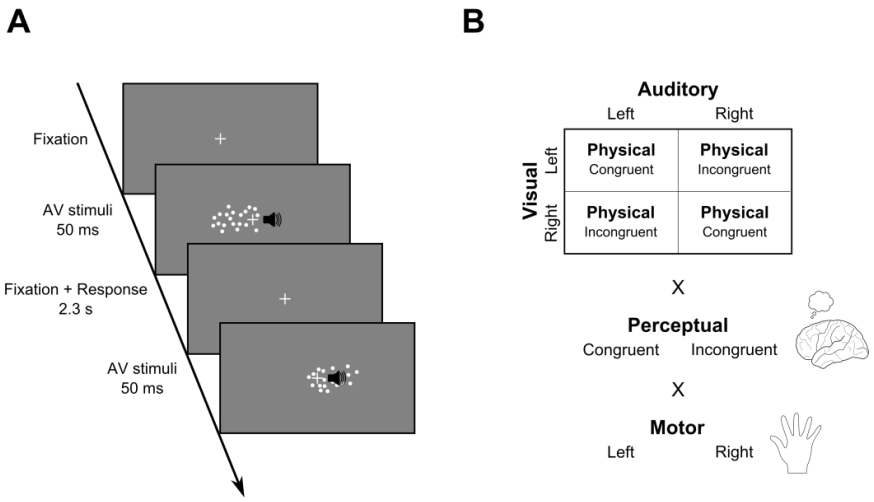


Figure 3.2 Experimental stimuli and design. (A) Time course of 2 example stimuli (the first physically incongruent, the second physically congruent). **(B)** Factorial design of the stimuli. Note that 3 factors (visual, auditory, motor) are independent variables manipulated by the experimenter, whilst perceptual congruency is a dependent variable on participants'

responses. Physical congruency is derived from the auditory and visual correspondences. Abbreviations: AV, audio-visual.

Eye movement recording and analysis

To address potential concerns that our results may be confounded by eye movements, we evaluated participants' eye movements based on eye tracking data recorded concurrently during the common source judgment task inside the mock scanner. Eye recordings were calibrated (~35° horizontally and ~14° vertically) to determine the deviation from the fixation cross. Fixation position was post-hoc offset corrected. For each position, the number of saccades (radial velocity threshold = 30°/s, acceleration threshold = 8000°/s^2, motion threshold = 0.15°, radial amplitude > 1°) and eye blinks were quantified (0-875 ms after stimulus onset). Post-stimulus saccades were detected in 33.5% ± 10.6% (mean ± SEM) of the trials. Critically, the 2 (visual left, right) x 2 (auditory left, right) repeated measures ANOVAs on the stimulus conditions performed separately for (i) % saccades or (ii) % eye blinks revealed no significant main effects or interactions indicating that differences between conditions are unlikely to be due to eye movement confounds.

Experimental setup

Visual and auditory stimuli were presented using Psychtoolbox version 3.0.11 (Brainard, 1997; Kleiner, Brainard, & Pelli, 2007; Pelli, 1997) running under MATLAB R2011b (MathWorks Inc.) on a MacBook Pro (Mac OSX 10.6.8). For the main task, visual stimuli were back projected to a Plexiglas screen using a D-ILA projector (JVC DLA-SX21) visible to the subject through a mirror mounted on the magnetic resonance (MR) head coil. Auditory stimuli were delivered via Sennheiser HD 280 Pro (in the anechoic chamber), Sennheiser HD 219 (in the mock scanner) and MR Confon HP-VS03 headphones (in the scanner).

Participants' eye movements were recorded in the mock scanner using an Eyelink II system (SR Research Ltd.) at a sampling rate of 1000 Hz.

MRI data acquisition

A 3T Philips Achieva scanner was used to acquire both T1-weighted anatomical images (TR/TE/TI, 8.4/3.8/min. 540 ms; 175 slices; image matrix, 288 x 232; spatial resolution, 1 x 1 x 1 mm^3 voxels) and T2*-weighted echo-planar images (EPI) with blood oxygenation level-dependent (BOLD) contrast (fast field echo; TR/TE, 2600/40 ms; 38 axial slices acquired in ascending direction; image matrix, 80 x 80; spatial resolution, 3 x 3 x 3 mm^3 voxels without gap). There were 10-12 runs with 240 volumes per run typically over 2 sessions. The first 4 volumes were not acquired to allow T1 equilibration effects. In one participant, we repeated one session, because in the excluded session the participant's behavioural performance was 15% lower compared to the mean accuracy of the remaining sessions. In another participant, 2 runs were excluded because of technical problems detected after data acquisition. In four participants, one or two runs were excluded to counterbalance the left vs. right response hand across runs.

fMRI analysis: Data pre-processing

The data were analysed with statistical parametric mapping (SPM8; Wellcome Trust Centre for Neuroimaging, London, UK; http://www.fil.ion.ucl.ac.uk/spm/; Friston, Holmes, Worsley, et al., 1995) running on MATLAB R2014a. Scans from each participant were realigned using the first as a reference, unwarped and corrected for slice timing. The time series in each voxel were high-pass filtered to 1/128 Hz. For the conventional univariate analysis, the EPI images were spatially normalized into MNI standard space (Ashburner & Friston, 2005), resampled to 2 x 2 x 2 mm^3 voxels, and spatially smoothed with a Gaussian kernel of 6 mm FWHM. For the multivariate decoding analysis, the EPI images were analysed in native participant space

and spatially smoothed with a Gaussian kernel of 3 mm FWHM. For the retinotopic analysis, the data were analysed in native space and without additional smoothing.

fMRI data analysis: Main experiment

Data were modelled in an event-related fashion with regressors entered into the design matrix after convolving each event-related unit impulse (representing a single trial) with a canonical hemodynamic response function and its first temporal derivative. Realignment parameters were included as nuisance covariates to account for residual motion artefacts.

Univariate fMRI analysis: For the conventional univariate analysis, the general linear model linear model modelled the 16 conditions in our 2 visual (left, right) x 2 auditory (left, right) x 2 perceptual congruency (congruent, incongruent) x 2 hand (left, right) factorial design. Condition-specific effects for each subject were estimated according to the general linear model (GLM) and passed to a second-level repeated measures ANOVA as contrasts. Inferences were made at the second level to allow for random effects analysis and inferences at the population level (Friston, Holmes, Price, Büchel, & Worsley, 1999). At the second between-subjects level we tested for the effects of visual signal location, auditory signal location, hand response (left vs. right), physical congruency (i.e. A and V signals coming from same or separate sources), and perceptual congruency (i.e. perceived as coming from same or separate sources).

We report activations at $p < 0.05$, corrected at the cluster level for multiple comparisons within the entire brain using an auxiliary uncorrected voxel threshold of $p < 0.001$ (Friston, Worsley, Frackowiak, Mazziotta, & Evans, 1994). For visualization, results of the random effects analysis were superimposed onto a single subject template image.

Multivariate decoding analysis: To allow for unbiased multivariate decoding results it is critical that parameter estimates are based on the same number of trials. Yet, when

conditions are defined based on participants' choice, the number of trials will differ across conditions. In order to ensure that our MVPA is nevertheless unbiased, we therefore generated design matrices for each condition modelled by up to 7 regressors, each based on exactly 8 trials. As a result of this subsampling procedure, each parameter estimate entering cross-validation was estimated with comparable reliability. After this subsampling procedure, the remaining trials (modulus) of all conditions were entered into a separate regressor. To ensure the decoding results do not depend on particular subsamples we repeated this subsampling procedure (with deriving GLM estimation and MVPA) 10 times.

Importantly, to dissociate perceptual from physical congruency, visual or auditory location or motor response, we ensured that the parameter estimates pertaining to perceptually congruent and incongruent conditions were matched with respect to all the other factors as auditory, visual, physical congruency and motor responses. This allowed us to identify regions encoding participants' causal judgment unconfounded by bottom-up physical congruency, auditory or visual location or motor output. Likewise, we decoded participant's motor response unconfounded by auditory or visual location, perceptual or physical congruency.

For decoding, we trained a linear support vector classification model as implemented in LIBSVM 3.20 (Chang & Lin, 2011). More specifically, the voxel response patterns were extracted in a particular region of interest (e.g. A1, see below for definition of region of interest) from the parameter estimate images corresponding to the magnitude of the BOLD response for each condition and run from one of the three GLMs as described above. To implement a leave-one-run-out cross-validation procedure, parameter estimate images from all but one run were assigned to the training data set and images from the 'left-out run' were assigned to the test set. Parameter estimate images for training and test data set were

normalized independently using Euclidean normalization of the images and mean centering of the features. Support vector classification models were trained to learn the mapping from the condition-specific fMRI responses patterns to the class labels from all but one run according to the following dimensions: (i) visual decoding (visual left vs. right); (ii) auditory decoding (auditory left vs. right); (iii) physical congruency decoding (physical congruent vs. incongruent); (iv) perceptual congruency decoding (perceptual congruent vs. incongruent); and (v) motor response decoding (hand left vs. right). The model then used this learnt mapping to decode the class labels from the voxel response patterns of the remaining run.

Non-parametric statistical inference was performed at the second, random-effects (i.e. 'between-subjects') level to allow for generalization to the population (Nichols & Holmes, 2002). First, we permuted the condition-specific labels of the parameter estimates for each run (to respect run-specific auto-correlations) and subject to determine chance decoding accuracy individually for each subject as the average decoding accuracy across all permutations (500 per GLM subsampling x 10 = 5000 permutations). At the 2^{nd} 'between-subjects' level we generated a null distribution of decoding accuracy by randomly assigning +/- sign to the subject-specific differences of observed minus chance decoding accuracy and repeating this procedure for all possible sign assignments (2^{10} = 1024 cases for 10 participants). Unless otherwise stated, results are reported at $p < 0.05$ (based on one sided tests). P values are Bonferroni corrected for multiple comparisons across all regions of interest unless we test the following a priori hypotheses: visual left/right location in V1, V2, V3, V3AB; auditory left/right location in A1, PT; motor left/right hand response in M1 and perceptual congruent/incongruent in DLPFC.

Visual retinotopic localizer

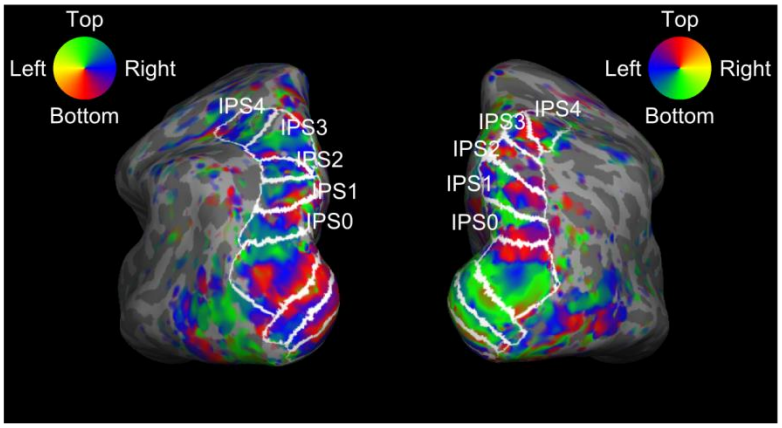
Standard phase-encoded polar angle retinotopic mapping (Sereno et al., 1995) was used to define visual and parietal regions of interest (Rohe & Noppeney, 2015a). Participants viewed a checkerboard background flickering at 7.5 Hz through a rotating wedge aperture of 70° width. The periodicity of the apertures was 44.2 s. After the fMRI pre-processing steps (see fMRI analysis: data pre-processing), visual responses were modelled by entering a sine and cosine convolved with the hemodynamic response function as regressors in a general linear model. The preferred polar angle was determined as the phase lag for each voxel, which is the angle between the parameter estimates for the sine and the cosine. The preferred phase lags for each voxel were projected on the participants' reconstructed and inflated cortical surface using Freesurfer 5.3.0 (Dale, Fischl, & Sereno, 1999). Visual regions V1–V3, V3AB, and parietal regions IPS0-IPS4 were defined as phase reversal in angular retinotopic maps. IPS0–4 were defined as contiguous, approximately rectangular regions based on phase reversals along the anatomical IPS (Swisher, Halko, Merabet, McMains, & Somers, 2007) and guided by group-level retinotopic probabilistic maps (Wang, Mruczek, Arcaro, & Kastner, 2015). See Figure 3.3 for example retinotopic delineation in a participant.

Region of interests used for decoding analysis

For the decoding analyses, all regions of interest (ROI) were combined from the left and right hemispheres.

Occipital, parietal and FEF regions: Regions in the occipital and parietal cortices were defined based on retinotopic mapping (see above). The frontal eye-field (FEF) was defined by an inverse normalized group-level retinotopic probabilistic map (Wang et al., 2015). The resulting subject-level probabilistic map was thresholded at 80 percentile and any overlaps with the motor cortex (for definition, see below) were removed.

Auditory, motor and prefrontal regions: The remaining regions were based on labels of the Destrieux atlas of Freesurfer 5.3.0 (Dale et al., 1999; Destrieux, Fischl, Dale, & Halgren, 2010). The primary auditory cortex region was defined as the anterior transverse temporal gyrus (Heschl's gyrus). The higher auditory cortex region was formed by merging the transverse temporal sulcus and the planum temporale (PT). The motor cortex region was based on the precentral gyrus. The dorsolateral prefrontal cortex (DLPFC) was defined by combining the superior and middle frontal gyri and sulci as previously described (Yendiki et al., 2010). In line with (Rajkowska & Goldman-Rakic, 1995) we limited the superior frontal gyrus and sulcus to Talairach coordinates y=26 and 53 and the middle frontal gyrus and sulcus to Talairach coordinates y=20 and 50.


Top
Left
Right
Bottom
IPS4
IPS3
IPS2
IPS1
IPS0
Top
Left
Right
Bottom
IPS4
IPS3
IPS2
IPS1
IPS0

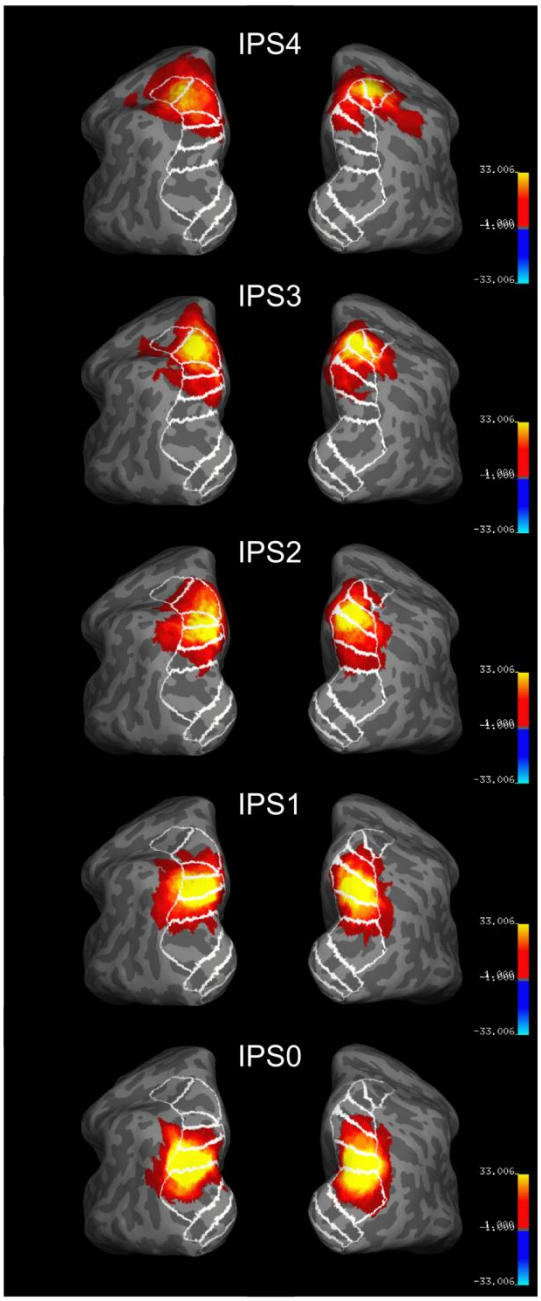

IPS4
IPS3
IPS2
IPS1
IPS0

Figure 3.3 Posterior view of example retinotopic delineation in a participant. Top: Delineations of visual and parietal regions based on phase reversals in the left and right hemispheres with insets of colour wheel. Functional activations are overlaid on inflated cortical surfaces with intraparietal sulcus (IPS0-4) labelled specifically. Field of view orientations are indicated in both colour wheels. **Bottom:** Group-level retinotopic probabilistic maps of the intraparietal sulcus (IPS0-4) from Wang et al (Wang et al., 2015) inverse-normalised to the participant's anatomical space and overlaid on the same cortical surfaces. The colour bar insets represent probabilities across subjects.

Results

Behavioural results

Firstly, we analysed the accuracy performance of participants. We adjusted the audio-visual disparities to each individual in order to get a threshold performance in their common source judgements. Indeed, behavioural results in the scanner confirmed that participants were at threshold when deciding whether auditory and visual signals were caused by common or independent events (Figure 3.4) with a small bias towards common source judgments (accuracy: 71% ± 0.20%; d-prime: 1.07 ± 0.12; bias: 0.16 ± 0.03; mean ± SEM in all cases).

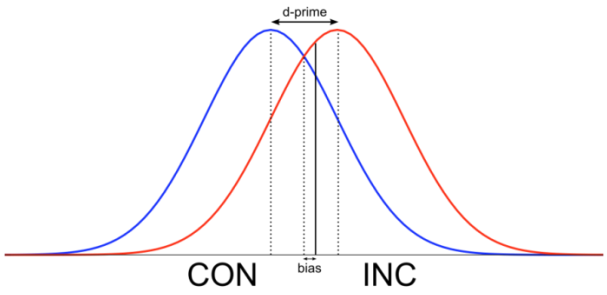


Figure 3.4 Signal detection analysis of the behavioural common source judgement results inside the scanner. Participants were at threshold when discriminating between

physically congruent (CON) and incongruent (INC) audio-visual signals illustrated by the highly overlapping distributions. D-prime is defined as the distance between distributions. Bias is the deviation from an ideal criterion set at the intersection of the two distributions.

Next, we pooled over the visual and auditory locations and performed a 2 (physical congruent, incongruent) x 2 (perceptual congruent, incongruent) repeated measures ANOVA on participants' response times (for descriptive statistics, see Table 3.1). Our choice to test this 2 x 2 design is justified by the taken MVPA analysis approach and the need to control whether RTs can confound our decoding results. A significant main effect of perceptual congruency ($F(1,9)=8.266$, $p=0.018$) and a significant physical congruency x perceptual congruency interaction was revealed ($F(1,9)=15.621$, $p=0.003$). Participants were slower on trials where they perceived audio-visual signals as caused by different events. Post hoc paired t-tests of the simple main effects revealed that participants were significantly faster judging congruent stimuli as congruent and incongruent stimuli as incongruent. In other words, they were faster on their correct than wrong responses suggesting that trials with wrong responses were associated with a greater degree of perceptual uncertainty. Crucially, this significant interaction cannot bias our decoding results, since correct and incorrect responses are evenly distributed in all of our stimulus classes due to our matching procedure described in the Multivariate decoding analysis subsection of the methods section.

Perceptual **Physical**	**Congruent**	**Incongruent**
Congruent	0.89 s (±0.05)	1.02 s (±0.06)
Incongruent	0.96 s (±0.06)	0.93 s (±0.06)

Table 3.1 Group-level reaction times for 2 (physical congruent, incongruent) x 2 (perceptual congruent, incongruent) design (SEM in parentheses).

fMRI analysis: univariate results

The current study focuses primarily on multivariate pattern analyses to characterize explicit causal inference in audio-visual perception. However, for the sake of completeness, we also give a short summary of conventional univariate analyses (see Table 3.2, Figure 3.5 and Figure 3.6).

Main effects of auditory and visual stimulus location

As expected, we found lateralization effects for visual and auditory stimuli. Right relative to left visual stimuli increased activations in the left middle and superior occipital gyri, whilst left relative to right visual stimuli increased activations in the right lingual gyri. Similarly, right relative to left auditory stimuli increased activations in the left planum temporale.

Main effect for left vs. right motor response

Left relative to right hand motor responses increased activations in the right pre- and postcentral gyri associated with sensory-motor processing.

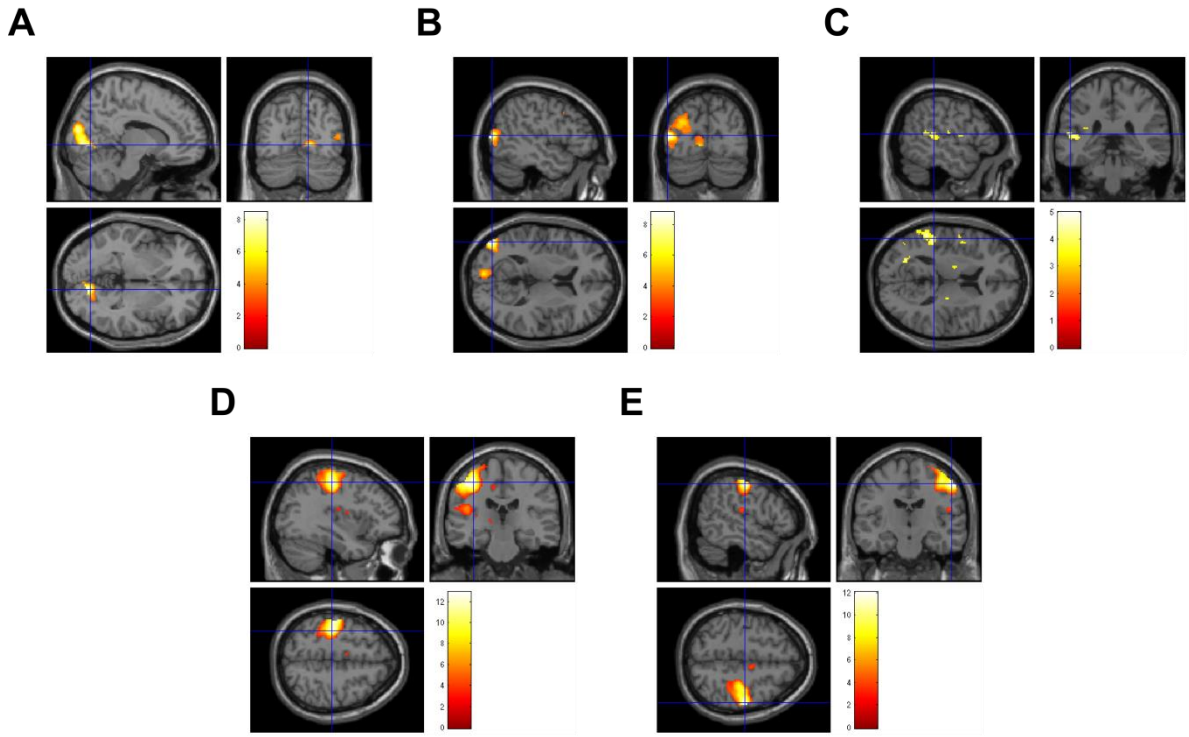


Figure 3.5 Classical univariate results of the sensory main effects. (A) Main effect of visual L > visual R. **(B)** Main effect of visual R > visual L. **(C)** Main effect of auditory R > auditory L. **(D)** Main effect of motor R > motor L. **(E)** Main effect of motor L > motor R. Activations are reported at $p < 0.05$, corrected at the cluster level for multiple comparisons within the entire brain using an auxiliary uncorrected voxel threshold of $p < 0.001$.

Main effect of physical and perceptual congruence

We did not observe any significant effects of physical congruency (i.e. interaction between visual and auditory location), however, a widespread right lateralized system including the frontal, insular, parietal and occipital cortices showed increased activations for perceptually incongruent relative to congruent stimuli.

Interaction between physical and perceptual congruency

To understand the interaction between physical and perceptual congruency, we note that the interaction is equivalent to correct vs. incorrect responses. We found bilateral putamen

activations for correct responses that is in concordance with previous results showing a role of putamen in overlearned tasks (von Saldern & Noppeney, 2013). For incorrect responses, we observed increased activations in bilateral prefrontal cortices and insulae that have been associated previously with greater executive demands (Noppeney, Josephs, Hocking, Price, & Friston, 2008; Werner & Noppeney, 2010).

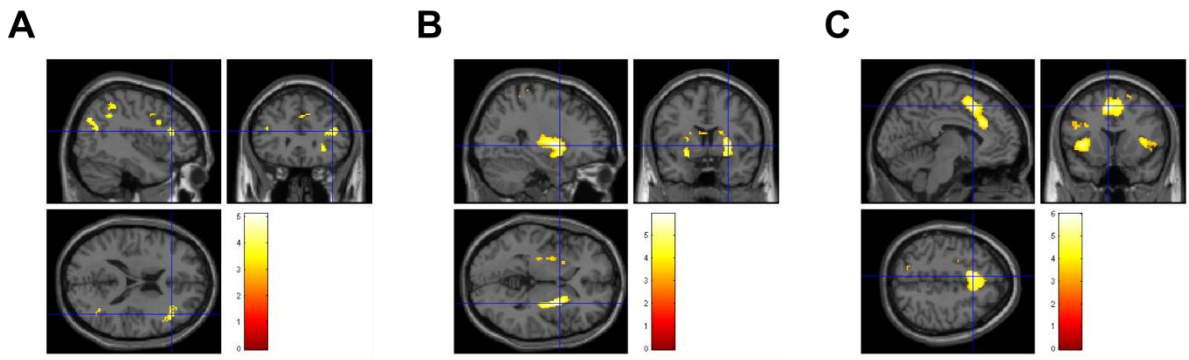


Figure 3.6 Classical univariate results of the main effects of perceptual congruency and the interaction between perceptual and physical congruency. (A) Main effect of perceptual incongruent > perceptual congruent. **(B)** Interaction of perceptual congruent x physical congruent (correct > incorrect) **(C)** Interaction of perceptual congruent x physical congruent (incorrect > correct). Activations are reported at $p < 0.05$, corrected at the cluster level for multiple comparisons within the entire brain using an auxiliary uncorrected voxel threshold of $p < 0.001$.

Region	MNI coordinates (x y z)			z-score (peak)	p-value (cluster)	Number of voxels
visual L > visual R						
R. lingual gyrus	12	-72	-2	7.58	<0.001	935
R. cuneus	10	-86	20	7.04		
R. lingual gyrus	10	-82	0	6.92		
visual R > visual L						
L. middle occipital gyrus	-48	-78	10	7.80	<0.001	1869
L. middle occipital gyrus	-20	-86	20	6.96		
L. superior occipital gyrus	-10	-86	2	5.87		
auditory R > auditory L						
L. Planum Temporale	-56	-44	14	4.66	<0.001	274
L. Planum Temporale	-52	-34	12	4.47		
L. Planum Temporale	-44	-42	16	3.92		
perc incongruent > perc congruent						
R. middle occipital gyrus	38	-62	22	4.10	<0.001	229
R. middle occipital gyrus	40	-74	34	3.87		
R. middle occipital gyrus	30	-62	38	3.75		
R. inferior parietal lobule	46	-42	52	3.72	0.001	183
R. inferior parietal lobule	38	-46	38	3.50		
R. inferior parietal lobule	40	-46	48	3.49		
R. middle frontal gyrus	42	30	18	4.09	0.002	179
R. inferior frontal gyrus (p triangularis)	50	20	8	4.08		
R. middle frontal gyrus	34	34	16	3.54		
R. middle frontal gyrus	26	6	52	3.89	0.004	150
R. superior frontal gyrus	24	18	50	3.18		
R. insula (anterior)	30	26	-6	4.86	0.006	139
R. precuneus	4	-68	46	3.73	0.018	112

Table 3.2 fMRI univariate results (L, left; R, right; perc, perceptual; phys, physical).

Region	MNI coordinates (x y z)			z-score (peak)	p-value (cluster)	Number of voxels
motor L > motor R						
R. postcentral gyrus	54	-16	50	65535	<0.001	1964
R. precentral gyrus	40	-16	54	65535		
R. postcentral gyrus	36	-36	52	4.16		
motor R > motor L						
L. precentral gyrus	-36	-24	52	65535	<0.001	2153
L. postcentral gyrus	-44	-24	50	65535		
L. postcentral gyrus	-52	-18	50	65535		
L. Rolandic operculum	-46	-22	18	6.04	<0.001	346
L. Rolandic operculum	-38	-16	20	3.48		
interaction of perc congruent x phys congruent (correct > incorrect)						
R. putamen	28	6	0	5.60	<0.001	757
R. putamen	30	-6	2	5.52		
R. putamen	26	6	-10	4.92		
L. putamen	-26	2	-8	4.73	<0.001	388
L. putamen	-24	10	-4	4.44		
L. putamen	-26	-2	8	4.38		
interaction of perc congruent x phys congruent (incorrect > correct)						
L. medial frontal gyrus (posterior)	-6	14	50	5.63	<0.001	1589
R. medial frontal gyrus (posterior)	6	12	54	5.12		
L. midcingulate cortex	-2	20	38	4.89		
L. inferior frontal gyrus (p triangularis)	-50	22	26	5.32	<0.001	716
L. insula	-36	18	6	5.47	<0.001	585
R. insula	38	16	6	4.27	<0.001	217
R. inferior frontal gyrus (p opercularis)	50	18	0	3.71		

Table 3.2 fMRI univariate results (continued)

fMRI analysis: multivariate results

We used multivariate pattern analyses to investigate the brain regions that encode: (i) visual location (left vs. right); (ii) auditory location (left vs. right); (iii) physical congruency (congruent vs. incongruent); (iv) perceptual congruency (congruent vs. incongruent); and (v) motor response (left vs. right hand). See Figure 3.7 for further details.

Decoding of auditory and visual location

Visual location was primarily encoded in visual areas including V1, V2, V3 and V3AB. In addition, visual information was maintained in the parietal cortex (IPS0-4) as well as in the frontal eye fields consistently with the well-established retinotopic organization of those cortical regions.

Similarly, auditory location could be decoded from the higher auditory cortical area, planum temporale. Propagated auditory information could be decoded then from the posterior parietal cortex (IPS0-2), the frontal eye fields, and finally the dorsolateral prefrontal cortex (DLPFC).

Decoding of physical and perceptual congruency

Physical congruency could be decoded from the parietal cortex (IPS0-4), also known as a classical multisensory convergence zone. Although to a lesser extent than auditory location, physical congruency was present in planum temporale. Finally, the FEF and DLPFC were also involved in encoding physical congruency.

Altogether, these results so far are consistent with the classical view of multisensory processing. Namely, that lower level sensory cortices encode primarily one sensory modality e.g. A1 auditory signals and V1 visual signals, then these sensory streams converge in higher order and association cortices. The significant decoding of physical congruency besides the maintained visual and auditory information in the higher order cortices confirms that indeed,

there is interaction between these sensory signals in planum temporale, parietal and frontal areas.

Critically, our experimental design allowed us to identify regions encoding perceptual congruency, i.e. participants' common source judgements irrespective of the physical congruency of the audio-visual signals. In line with our predictions, participants' perceived congruency could be decoded from DLPFC. Moreover, perceptual congruency could be better decoded from DLPFC than any other stimulus feature that suggests a key role of DLPFC in causal inference. Interestingly though, perceptual congruency could be decoded to a lesser extent in the widespread system of FEF, IPS0-4 and even at the early visual processing stage of V2.

Decoding of motor response

We ensured that participants' reports on perceived congruency was orthogonal to their motor response by alternating the mapping from participants' common source judgment to selected hand response across runs. Not surprisingly, the motor response was extremely well decoded from the precentral (and postcentral) gyrus. However, the motor response was not selectively encoded in the sensory-motor cortex, but in many other tested ROIs including the FEF, IPS0-4 and V3AB. Much more surprisingly, we were also able to decode participants' motor response from planum temporale and Heschl's gyrus. We suspect that decoding sensory-motor information from Heschl's gyrus might be artefactual due to e.g. smoothing activations from the neighbouring secondary somatosensory areas.

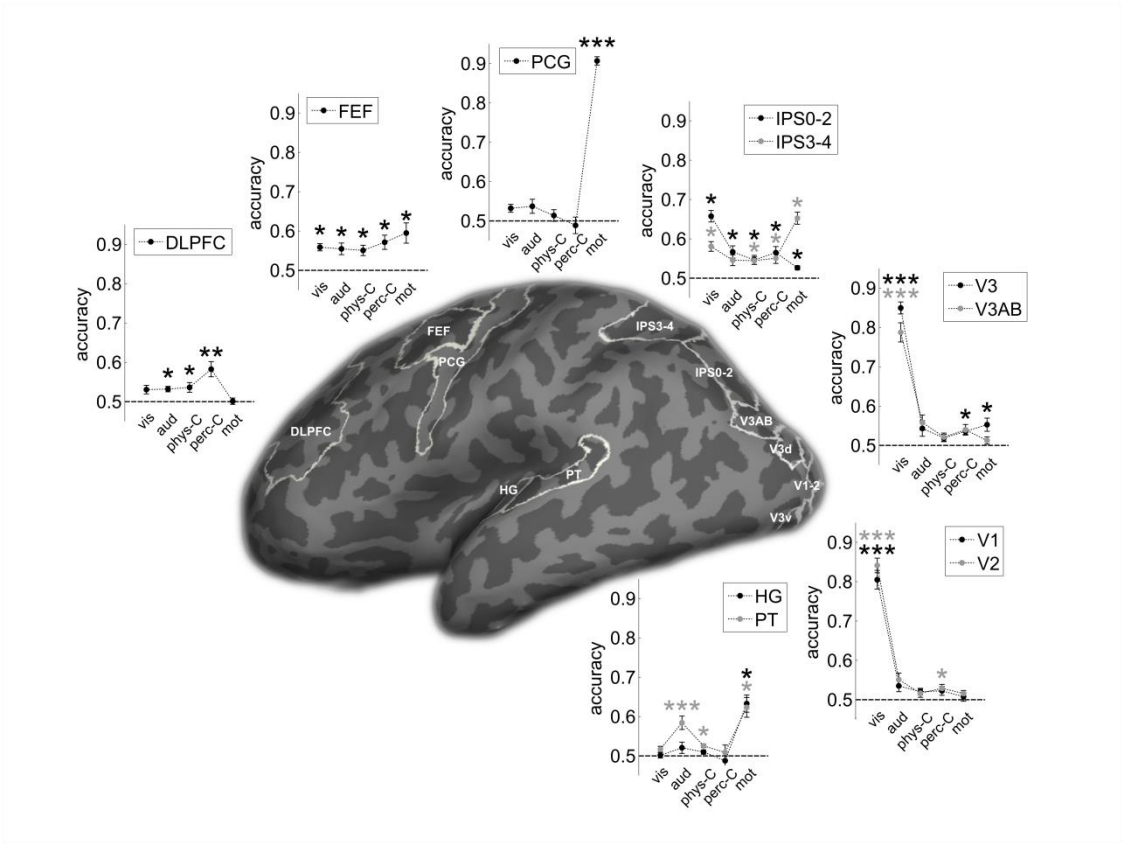


Figure 3.7 Multivariate pattern results along the visual and auditory spatial cortical hierarchy. Decoding accuracy and significance (* = p<0.05, ** = p<0.005, *** = p<0.001 Bonferroni corrected for multiple comparisons unless a priori hypotheses were tested) were tested via classification of: (i) vis = visual left vs. right location (ii) aud = auditory left vs. right location, (iii) phys-C = physical congruency vs. incongruency, (iv) perc-C = perceptual congruency vs. incongruency, and (v) mot = motor left vs. right hand response in the regions of interest (ROI) as indicated in the figure. Group mean and SEM values are marked by circles and error bars. The regions of interest are delineated on the surface of a single subject brain. Abbreviations for *ROI*: primary, secondary and higher order visual regions, V1, V2 and V3, V3AB, respectively; Heschl's gyrus, HG; planum temporale, PT; intraparietal sulcus, IPS0-2 and IPS3-4; precentral gyrus, PCG; frontal eyefields, FEF; dorsolateral prefrontal cortex, DLPFC.

Discussion

To make a coherent perception of the world the brain needs to integrate sensory signals generated by the same event and segregate those generated by different events (Trommershäuser et al., 2011). The human brain infers whether or not signals are originating from a same source or event based on multiple correspondence cues such as spatial disparity, temporal synchrony and semantic or higher order congruency (Chen & Vroomen, 2013; Laurienti et al., 2004; HweeLing Lee & Noppeney, 2011; Recanzone, 2009; Slutsky & Recanzone, 2001; Welch, 1999). For instance, human observers are more likely to infer that auditory and visual signals originate from different sources, when audio-visual spatial discrepancy increases. Therefore, observers' causal inference judgments are inherently correlated with the spatial correspondences of the audio-visual signals making it challenging to dissociate causal inference judgment (i.e. perceived congruency) from physical congruency in standard experiments.

To dissociate participants' causal judgment from the physical bottom-up congruency cues of the signals we followed two critical steps. Firstly, we adjusted the audio-visual spatial disparity individually for each participant. Spatially congruent audio-visual signals were perceived as coming from the same source in ~70% of cases. Similarly, spatially disparate audio-visual signals were perceived as coming from independent sources in ~70% of cases. Secondly, this causal uncertainty of the signals allowed us to select and compare physically identical audio-visual signals being perceived as either coming from common or independent sources. In addition, to dissociate participants' causal judgment from their motor responses we counterbalanced the mapping from participants' causal judgment to their response selection over runs. In summary, our experimental design enabled us to characterize a system of brain regions in audio-visual spatial processing with respect to five different

representations: (i) visual space (left vs. right); (ii) auditory space (left vs. right); (iii) physical spatial congruency; (iv) perceptual congruency (congruent vs. incongruent); and (v) motor response (left vs. right hand).

Unsurprisingly, our multivariate decoding results demonstrate that low level visual areas (V1-3) encode predominantly visual space, the planum temporale (PT) auditory space and precentral gyrus participant's motor responses. Further, physical congruency could be decoded from higher order visual or auditory regions (IPS0-4, planum temporale) and prefrontal cortices (DLPFC, FEF). These results are consistent with the classical views of a hierarchical organization of multisensory perception, where low level sensory cortices process sensory information from their preferred modalities, whilst multisensory information converges in higher order cortical regions (Felleman & Van Essen, 1991). In recent years this view has been challenged by the demonstration of multisensory interactions already at the primary cortical level (Driver & Noesselt, 2008; Ghazanfar & Schroeder, 2006). However, the majority of these interactions in primary sensory areas demonstrated cross- modal influence on the preferred sensory modality by the non-preferred sensory signal rather than a formal interaction of the two incoming signals (Kayser et al., 2007; Lakatos, Chen, O'Connell, Mills, & Schroeder, 2007; Werner & Noppeney, 2011). By contrast, the decoded physical congruency in planum temporale reflects the representational integration of the auditory and visual signals. Our results also corroborate previous results (Rohe & Noppeney, 2016) revealing integration of spatial information across the senses predominantly in higher order parietal, prefrontal areas and the planum temporale.

Importantly, our study also enabled us to identify regions encoding participants' causal judgment irrespective of bottom-up physical congruency cues. In line with our a priori predictions the DLPFC was the only region where the decoding accuracy profile peaked for

causal judgements (i.e. perceived congruency). This result indicates that the DLPFC encodes participants' explicit causal inference irrespective of the true physical audio-visual congruency or their motor response. Furthermore, DLPFC was not able to discriminate the motor response better than chance. Therefore, we think that the DLPFC accumulates the spatial sensory information into an explicit causal judgement leading to a final common or independent source decision.

Given the extensive evidence for early integration in early sensory cortices discussed above it is rather unlikely that the brain keeps from multisensory binding until an accumulated causal judgment made by DLPFC. On the contrary, it is more plausible that the brain integrates or segregates spatial sensory signals starting already at the primary cortex level and progressively refines the representations both over time and along the cortical hierarchy. Indeed, a recent study confirmed the hierarchical nature of spatial representations in the human brain (Rohe & Noppeney, 2015a).Therefore, the evidence about the world's causal structure that is accumulated in DLPFC needs to be projected backwards to lower level sensory areas to inform and update their spatial representation and the binding process. In line with such a feedback loop architecture we were able to decode perceptual congruency also from low level sensory cortices such as V2-3 and planum temporale suggesting that the causal inference in DLPFC top-down modulates along the sensory processing hierarchy.

Very interestingly, we were able to decode all the binary representations of our design from the frontal eye field (FEF) and the intraparietal sulcus (IPS0-4) including visual and auditory space, physical and perceptual congruency and motor responses. This decoding profile suggests that FEF and IPS form a circuitry, where they integrate audio-visual signal into spatial representations informed by the explicit causal inference encoded in DLPFC. Therefore our results extend previous findings showing that IPS3-4 arbitrates between audio-

visual integration and segregation and performs implicit causal inference depending on the bottom-up physical characteristics of the sensory signals (Rohe & Noppeney, 2015a). Despite their slightly different paradigms, these two studies collectively demonstrate that causal inference is a key factor in multisensory integration in posterior IPS (and FEF) to guide behavioural responses. Critically, our current paradigm also enabled us to orthogonalize participants' motor responses with respect to their causal judgments. Even when trials were matched for perceptual judgments we were able to decode participants' hand response from IPS0-4 significantly better than chance. Therefore, these results suggest that IPS0-4 integrate audiovisual signals not only into spatial representations, but it also transform them into motor mappings. Alternatively, motor response selection in premotor cortices may top-down modulate neural representations in IPS. In concordance with these findings, numerous electrophysiological studies have demonstrated that IPS can transform sensory input into motor output according to learnt mappings (Cohen & Andersen, 2004; Gottlieb & Snyder, 2010; Sereno & Huang, 2014).

In conclusion, our study was able to dissociate participants' causal inference from physical bottom-up congruency cues and motor responses. Our results suggest that DLPFC plays a key role in inferring the causal structure of audio-visual signals, which in turn influences audio-visual processing in FEF-IPS0-4 and lower sensory cortices as planum temporale and V2-3. Moreover, informed by the bottom-up congruency cues (i.e. physical congruency) and the inferred causal structure (i.e. perceptual congruency) FEF and IPS forms a circuitry to integrate auditory and visual spatial signals into representations that are transformed into motor response mappings.

CHAPTER 4: CHANGING THE TENDENCY TO INTEGRATE AUDIO-VISUAL SIGNALS

Introduction

The human brain receives signals typically in multiple sensory channels at the same time. To form a substantive representation of the world, the brain needs to decipher which signals belong to the same object and should be integrated, and which ones belong to different objects and hence should be kept separate (Adams, Graf, & Ernst, 2004; Körding et al., 2007; Magnotti, Ma, & Beauchamp, 2013; Roach et al., 2006; van Wassenhove, 2013). According to Bayesian models of perceptual inference the observer should infer the underlying causal structure of the world by combining bottom-up sensory correspondences with top-down prior beliefs and expectations (Körding et al., 2007; Rohe & Noppeney, 2015a, 2015b, 2016; Shams & Beierholm, 2010).

A vast body of research has demonstrated that the brain uses various bottom-up inter-sensory cues such as spatial collocation, temporal synchrony and semantic congruency to determine whether or not signals come from a common source and should be integrated into a unified percept (Adam & Noppeney, 2014; Chen & Vroomen, 2013; Körding et al., 2007; Hweeling Lee & Noppeney, 2014; Magnotti et al., 2013; Munhall, ten Hove, Brammer, & Paré, 2009; van Wassenhove, 2013; Wallace et al., 2004). Most prominently, spatial ventriloquism is known to decrease at large audio-visual spatial disparities or temporal asynchronies when it is unlikely for two signals to originate from a common source (Slutsky & Recanzone, 2001; Wallace et al., 2004). Yet, multisensory integration is not purely driven by bottom-up sensory correspondences; it also depends on top-down prior expectations that signals are generated by a common source. Multisensory integration is enhanced if task

instructions induce observers to believe that two signals come from a common source (Bedford, 2001; Helbig & Ernst, 2007; Warren et al., 1981; Welch & Warren, 1980). In real life, prior information about whether signals are likely to come from a common source are rarely conveyed via task-instructions or direct communication. Instead the brain needs to constantly update and adapt its so-called prior common source expectations based on the incoming sensory signals. For instance, when bombarded concurrently with many signals at busy crossroads, the brain should lower its common source expectations and tendency to bind signals into a coherent percept. By contrast, in a quiet two person conversation the brain should increase its prior common source expectations leading to greater multisensory integration. Indeed, for the McGurk illusion (McGurk & Macdonald, 1976) a series of recent studies have demonstrated that observers are more likely to integrate a visual 'ga' and an auditory 'ba' into an illusory 'da' percept (Gau & Noppeney, 2016; Nahorna, Berthommier, & Schwartz, 2012, 2015) when the McGurk trial is preceded by a series of audio-visually congruent phonemes. Critically, we have to note that similar phenomenon has already been described about 30-40 years ago in the unisensory domain, namely categorical perception of speech stimuli (Repp, 1984). All these findings suggest that observers dynamically adapt their prior common source expectations to changes in the environmental statistics thereby modulating their tendency to integrate signals into a coherent percept.

Yet, all of those experiments focused selectively on the McGurk illusion, which illustrates integration of higher order phonological information. By contrast, using the spatial ventriloquist illusion a recent study suggested that observers' tendency to integrate audio-visual signals into spatial representations is stable across days, but variable across observers (Odegaard & Shams, 2016). This raises the question, whether low level audiovisual integration processes as reflected in spatial ventriloquism are more automatic and immune to

changes in stimulus statistics than integration of high level audiovisual phonological information.

Using a spatial ventriloquist paradigm the current study presented observers with AV signals that were: (i) spatially collocated and synchronous; and (ii) spatially disparate and asynchronous. On each trial, participants reported the perceived sound location irrespective of the location of the visual signal. To manipulate observers' prior common source expectations we altered the relative frequencies of spatiotemporally congruent and incongruent trials over blocks. Based on Bayesian causal inference models we expected observers to be more likely to integrate AV signals as indexed by the ventriloquist illusion in blocks with a high probability of congruent AV trials.

Methods

Participants

Fifteen participants (14 females; 14 right-handed; age: mean=20.5, SD=3.4; age and handedness of 1 participant is unknown) with no history of neurological or psychiatric illness gave informed consent to take part in the main experiment. One participant was excluded after 1 completed run in the experiment, in which case the previous adaptive staircases did not converge within 28° audio-visual disparity 3 out of 4 cases. The same participant also turned to be an outlier with respect to the mean contextual effect based on the completed run (deviant with > 2SD). All participants reported normal hearing and normal or corrected-to-normal vision. The study was approved by the human research ethics committee of the University of Birmingham.

Experimental procedure

Participants completed the study in 2 sessions. The first session lasted for ~1 hour. It consisted of the recording of sound stimuli via assessing participants' sound localization

performance. Unbiased recordings were chosen and selected for the second session that lasted for ~1.5 hours. The second session was further divided up to 2 parts. In the first part, adaptive staircase runs were performed to determine subject-specific spatial disparities. These threshold disparities were then used in the main experiment. After the main experiment, questionnaire was filled out by the participants to assess their awareness of the experimental design and stimuli.

Stimuli

The visual stimulus was a cloud of 20 white dots (diameter: 0.4° visual angle) sampled from a bivariate Gaussian with horizontal and vertical standard deviations of 2° visual angle presented for 50ms on a grey background with 70% contrast. The sound stimuli were 50ms bursts of white-noise with 5ms ramp on and off. They were recorded for each subject in an anechoic chamber of the School of Psychology, University of Birmingham in a separate session prior to the experiment. The process consisted of displaying the sounds with an Apple Pro Speaker (at a distance of 68 cm from the participants) from -15° to 15° visual angle with 0.5° visual angle spacing along the azimuth, and recording with Sound Professionals™, Inc. (USA) in-ear binaural microphones. The participant's head was placed on a chin rest with forehead support and controlled by the experimenter to ensure stable positioning during the recording process. Five stimuli were recorded at each location to ensure that sound locations could not be determined based on irrelevant cues.

Assessment of sound localization performance

In a run, participants were presented with 15 stimulus locations from the recorded auditory signals (with smaller spacing around 0°) in a forced choice left-right discrimination task (10 repetitions per stimulus location). A cumulative Gaussian was fitted to the percentage 'perceived right responses' as a function of stimulus location using maximum-likelihood

estimation (www.palamedestoolbox.org) to obtain estimates of the threshold (point of subjective equality, PSE) and the slope (inverse of the standard deviation, STD) of the psychometric function. An initial recording and left-right discrimination task was performed to familiarize participants with the process. After that, multiple recordings and runs of left-right discrimination tasks followed each other until the PSE was smaller than ±1.5° using 20 repetitions per stimulus condition on the same recording. Runs with PSE larger than ±3° were excluded from the analysis due to possible head movement during recording or fatigue during the discrimination task.

Adaptive staircases to determine subject-specific spatial disparities in ventriloquist paradigm

To maximize the impact of contextual modulation, we adjusted the audiovisual spatial disparity individually for each participant such that the ventriloquist effect was of intermediate strength (see below). On each trial, auditory and visual stimuli, each of 50 ms duration were presented at a visual stimulus onset asynchrony of 2.4 sec. The audio-visual signals were temporally asynchronous with a fixed A leading asynchrony of 100 ms and spatially disparate with the disparity adjusted across trials in adaptive staircase. Auditory and visual stimuli were always presented in opposite hemifields equidistant from the central fixation cross. On each trial participants located the sound using a forced choice left/right key press. Using 1-up/2-down adaptive staircases (www.palamedestoolbox.org) we adjusted the audiovisual spatial disparity individually for each subject to allow for 70% auditory localization accuracy. After an initial practice run, participants completed four runs each including 40-80 trials. The adaptive staircase was terminated after 8 reversals within a window of 20 trials. Subject-specific spatial disparity thresholds were set to the mean audio-visual disparity across those final 20 trials and averaged across the four runs. The

convergence of staircases was re-assessed after the runs and staircases with poor convergence (causing a ≥45% decrease in SD of all relative to the remaining threshold estimates) were excluded from the analysis.

Main experimental design

The main experiment conformed to a 2 (AV spatiotemporally congruent vs. incongruent stimulus) x 2 (context: high vs. low frequency of spatiotemporally congruent stimuli). On each trial participants were presented with an audio-visual signal that was either spatially collocated and synchronous or spatially disparate and asynchronous (with A leading by 100 ms). In an auditory selective attention paradigm observers located the sound in a forced choice left vs. right response and constantly fixated a central cross. On half of the trials the auditory signal was presented in the left hemifield, on the other half of the trials it was presented in the right hemifield. The same applied to the visual signal. In the congruent context, audiovisual signals were (i) spatially collocated and temporally synchronous in 75% of the trials and (ii) spatially disparate and temporally asynchronous (i.e. A leading by 100 ms) in 25% of trials. In the incongruent context, the percentages of congruent and incongruent trials were reversed (Figure 4.1). The congruency context alternated in blocks of 32 trials. The initial context was counterbalanced across runs and subjects. Each participant completed 4 runs amounting to an overall 768 trials (1 participant completed only2 runs). Each run included 3-3 blocks of the two contexts. Participants were instructed that the visual and auditory signals could be generated by one common event or two independent events.

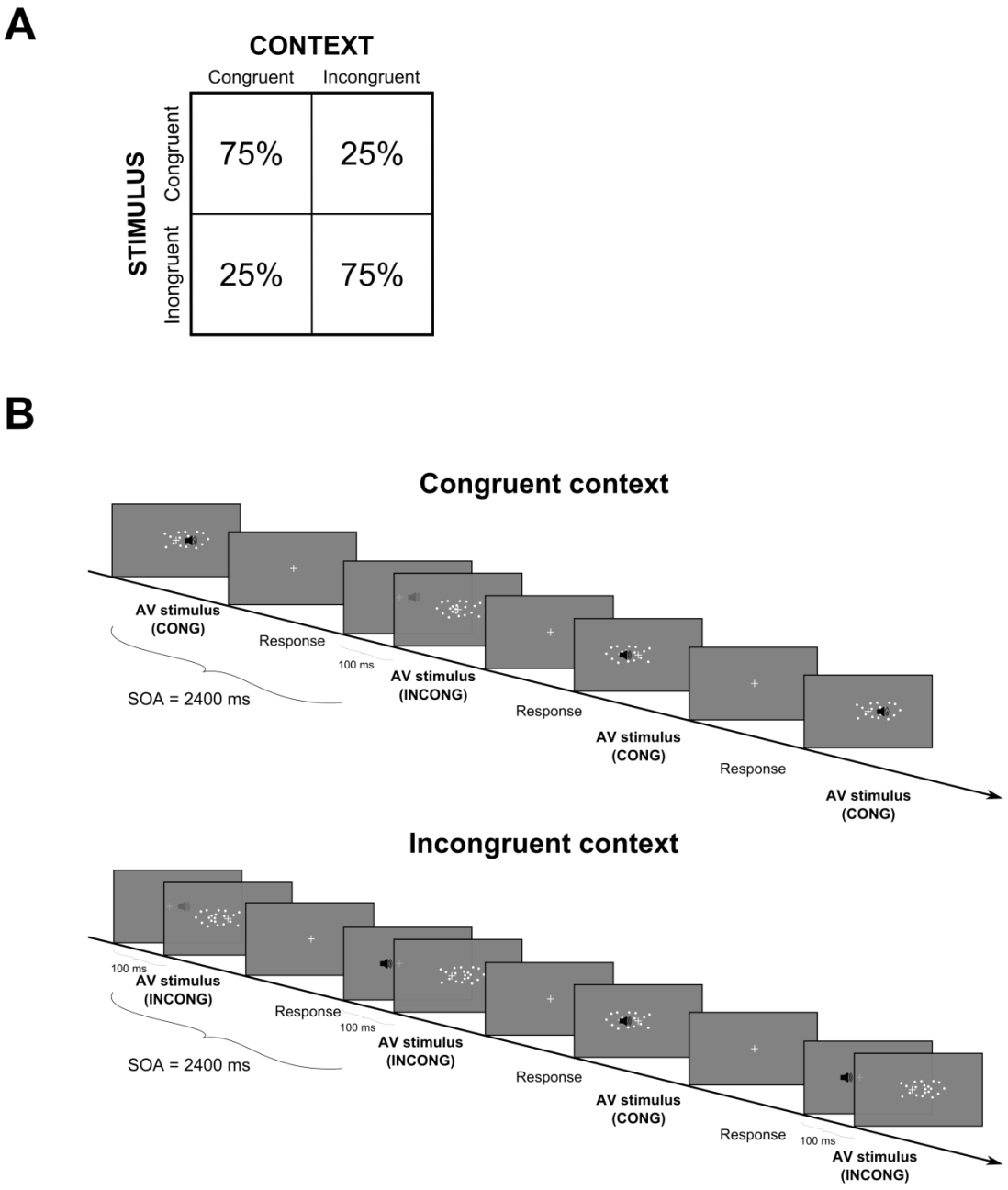


Figure 4.1 Experimental design and stimuli. (A) Factorial design of stimuli and context. **(B)** Time course of 4 example stimuli in both contexts. **Top:** Congruent context with 75% congruent (spatially collocated and synchronous) AV stimuli. **Bottom:** Incongruent context with 75% incongruent (spatially discrepant and asynchronous) AV stimuli. Abbreviations: SOA, stimulus onset asynchrony; AV, audio-visual; CONG, congruent; INCONG, incongruent.

Post-experimental questionnaire

To assess observers' awareness of experimental stimulus and contextual manipulations observers rated after the entire experiment: (i) the frequency of spatiotemporally congruent and incongruent audiovisual stimuli; and (ii) whether stimuli were blocked based on similar spatial or temporal properties (amongst other aspects in a questionnaire).

Experimental setup

Visual and auditory stimuli were presented using the Psychtoolbox version 3.0.11 (Brainard, 1997; Kleiner et al., 2007; Pelli, 1997) running under MATLAB R2011b (MathWorks) on a MacBook Pro (Mac OSX 10.6.8). Subjects were seated at a distance of 50 cm from a 24'' gamma-corrected LCD screen resting their head on a chinrest with forehead support. Sounds were delivered via circumaural headphones (Sennheiser HD 280 Pro).

Data analysis: Assessment of sound localization abilities

Cumulative Gaussian psychometric functions (PF) were fitted to participants' percentage of perceived right responses as a function of sound location to obtain estimates for the point of subjective equality as a measure of spatial bias, the slope and the just noticeable difference (JND) defined the difference between the abscissa values for 50% and 84% perceived right responses. A linear robust regression was estimated with participants' precision in unisensory sound localization (as quantified by JND of the auditory PF) as predictor and their spatial disparity threshold obtained from the ventriloquist adaptive staircases as dependent variable.

Data analysis: Main experiment

In the main experiment, the ventriloquist effect (VE) was calculated for each incongruent trial as:

$$VE = \frac{A_{resp} - A_{true}}{V_{true} - A_{true}}$$

where A_{true} and V_{true} indicated auditory and visual locations, and A_{resp} indicated the responded auditory location. We quantified the contextual modulation of VE in terms of the difference between the VE averaged across all incongruent trials in the congruent minus the incongruent context. To allow for generalization to the population we entered the contextual modulation for each subject into a one sample t-test.

Results

Relation between variability in sound localization and disparity threshold in VE

From the psychometric functions of the unisensory sound localization session, we obtained the STD as a measure of localization variability for each subject. Conversely, from the adaptive staircase session, we obtained the threshold spatial disparity that is associated with a ventriloquist illusion (i.e. AV integration) in 30% of the trials. A linear regression demonstrated that auditory STD as independent variable significantly predicts and observer's spatial disparity threshold as dependent variable ($t(12) = 3.059$, $p = 0.01$ for the slope parameter). In other words, the more precise participants are on sound localization, the smaller the spatial disparity required to attenuate the ventriloquist effect. This is consistent with a previous study demonstrating the auditory reliability determines the spatial audiovisual integration window (Rohe & Noppeney, 2015b).

Contextual modulation of the ventriloquist effect across blocks

In line with Bayesian theories of perceptual inference we observed a significant increase in the ventriloquist effect for blocks with 70% relative to 30% congruent trials. The effect was revealed by a one-sample t-test on the contextual modulation of the VE ($t(13) = 2.905$ $p=0.01$). Figure 4.2 illustrates the contextual modulation with additional descriptive statistics of all conditions for subjective as well as group-level indices. The group-level descriptive statistics can be found in Table 4.1. These results suggest that multisensory integration of

spatial signals does not only depend on low level inter-sensory correspondences but also on their top-down prior congruency expectations that rapidly adapt to changing environmental statistics. As in this experiment we manipulated audiovisual congruency concurrently in space (i.e. spatial disparity) and time (i.e. synchrony) we cannot dissociate whether either or both manipulations are effective for contextual modulation. Moreover, in this study we only included congruent and incongruent stimuli. Therefore an increase in the frequency of incongruent stimuli within a block may have facilitated processing of incongruent audio-visual stimuli thereby enabling observers to locate the sound of incongruent audio-visual signals more accurately.

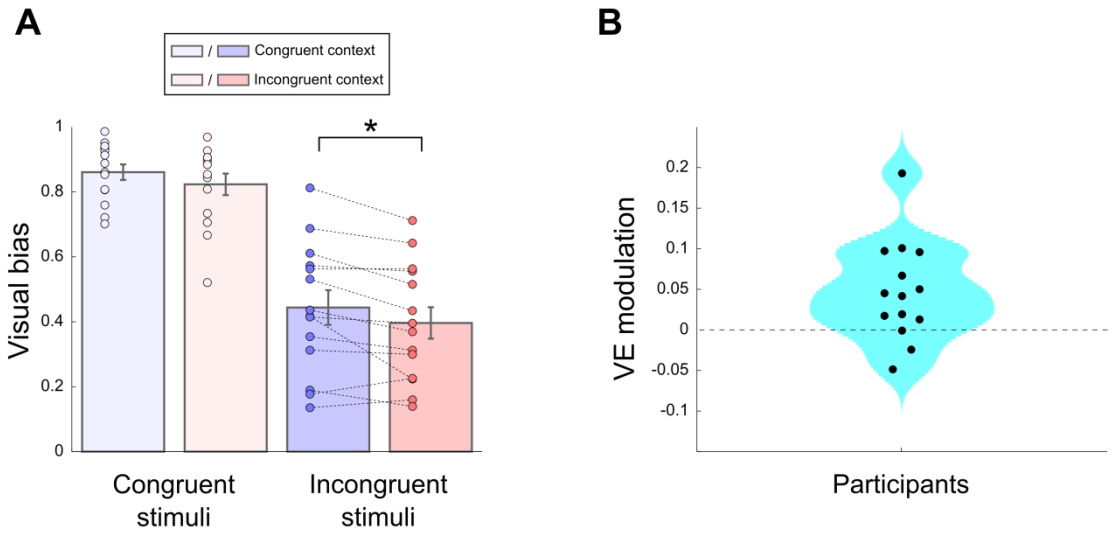


Figure 4.2 Contextual modulation of the ventriloquist effect (VE). (A) Visual bias as a function of context and stimuli. Bars represent group mean values, circles represent single subject results. Dotted lines between circles represent data of same participants for incongruent stimuli. Visual bias reflects proportion correct for congruent and VE for incongruent stimuli. **(B)** Violin plot of participants' data for contextual modulation of the VE calculated as VE in congruent context – VE in incongruent context.

CONTEXT STIMULI	Congruent	Incongruent
Congruent	0.86 (±0.02)	0.82 (±0.03)
Incongruent	0.44 (±0.05)	0.40 (±0.05)

Table 4.1 Group-level mean visual biases (SEM in parentheses).

Participants' awareness of contextual modulation

Our results demonstrated that higher frequency of congruent trials increases audiovisual binding and thereby the ventriloquist effect. We suggested that observers adapt their prior top-down congruency expectations that in turn shape multisensory integration. This raises the question whether observers explicitly adjust their prior common source expectations or whether these adaptation processes emerge automatically in the absence of observers' awareness. We assessed participants' awareness of changes in stimulus properties in a post-experimental questionnaire (1 participant did not fill in the questionnaire). Participants reported that 71.2% (±0.07% SEM) of the trials were synchronous, which is significantly different from the true 50% value (one-sample t-test, $t(12)= 2.856$, $p=0.01$) indicating that they were not fully aware of the synchrony manipulation. More importantly, only five participants (41.7%) reported that they noticed a periodical change in the frequency of trials with respect to the audio-visual spatial disparity (1 participant did not respond to this question) and only three participants (27.3%) reported a periodical change in the frequency of trials with respect to audio-visual asynchrony (2 participants did not respond to this question). The other participants reported that the stimuli were presented completely randomly with respect to the spatial and temporal characteristics of the signals.

Discussion

To establish a veridical representation of the world the brain needs to bind sensory signals generated by a single event and segregate those generated by multiple events. Therefore, multisensory perception is essentially linked to causal inference, i.e. the brain needs to decipher the underlying causal structure of the sources that generate the multisensory signals (Körding et al., 2007; Rohe & Noppeney, 2015a, 2015b, 2016; Shams & Beierholm, 2010). Bayesian modelling formulates that the solution of this so-called causal inference problem is achieved by combining prior knowledge that the signals are generated by a common source with the ongoing inter-sensory correspondences of the signals, such as spatiotemporal characteristics (Körding et al., 2007; Rohe & Noppeney, 2015b; Wozny et al., 2010). Numerous studies using such Bayesian modelling have demonstrated in recent years that human observers update their priors dynamically (Berniker, Voss, & Körding, 2010; Sato & Körding, 2014). In the current study, we investigated whether observers dynamically adapt their so-called 'common source prior' i.e. their prior belief about spatial signals coming from a common source to changes in the relative frequencies of audio-visual signals generated from common or separate sources. We used the spatial ventriloquist illusion as a prime example of audio-visual integration.

We manipulated the relative frequencies of trials where (i) auditory and visual signals were collocated and synchronous or (ii) spatially disparate and asynchronous in short blocks of 32 trials. We demonstrate that the ventriloquist effect occurs more often in blocks with a high frequency of spatiotemporally congruent trials. This finding suggests that observers dynamically adapt their common source expectations to the probability of congruent trials, that in turn results in an increased binding tendency and more frequently experienced ventriloquist illusion.

Our results are concordant with a previous study from Van Wanrooij and colleagues (Van Wanrooij et al., 2010) showing reduced response times in a ventriloquist paradigm when the probability of congruent trials was increased. Yet, this study did not show an effect on the audio-visual bias. We think that two reasons underlie the discrepancy. Firstly, in our experiment we adjusted the spatiotemporal characteristics of the audio-visual stimuli to maximize the possibility of contextual modulation. Namely, the spatial audio-visual disparities were adjusted to each participant to achieve an optimal ventriloquist effect, as well as temporal asynchrony of the signals being modulated within a hardly detectable time window to assist contextual modulations automatically. Secondly, changes in the probability of spatially disparate trials can potentially induce several counteracting effects. Firstly, in blocks with high probability of spatially collocated trials participants regularly receive a teaching signal for sound localization. These congruent teaching trials could enhance auditory localization accuracy, which would in turn reduce the ventriloquist effect on trials when the two stimuli are spatially disparate. Secondly, on the other hand, the more conflicting the audio-visual signals are, the higher likelihood that the task demand that might lead to more localization errors and increased ventriloquist effect in blocks of high probability of spatially disparate trials. This task difficulty confound might be more apparent when participants' selective attention is reduced (e.g. due to fatigue) or when they have some response bias to the visual signals. Thirdly, a recent study showed that spatial recalibration of auditory signals can occur after a single exposure to discrepant audio-visual signals. A higher probability of spatially disparate audio-visual signals might lead to higher spatial recalibration, and again leading to an increase in ventriloquist effect in the incongruent context. Importantly, the impact of these counteracting effects may depend on the exact relative frequencies and the

length of the contextual blocks. For instance, while we alternated the context every 32 trials, Van Wanrooij and colleagues compared contextual effects across sessions of 300-800 trials.

Our results demonstrate that human observers adapt their binding tendencies dynamically to changes in environmental statistics. Increased frequency of trials with spatially collocated and synchronous stimuli increased the occurrence of the spatial ventriloquist illusion. Interestingly, these modulatory effects were observed irrespective of whether observers were aware of the changes in stimulus statistics based on a post-experimental questionnaire. Therefore, our results corroborate previous research showing an increase in the emergence of the McGurk illusion when preceded by a series of synchronous audiovisual movies (Gau & Noppeney, 2016; Nahorna et al., 2012, 2015). Collectively, these studies demonstrate that dynamic updating of common source priors is a generic mechanism critical for multisensory integration. In other words, the brain uses the recent past to predict whether future signals are likely to come from a common source and should be integrated into a unified percept.

Future studies will need to investigate the cognitive and neural mechanisms underlying the dynamic changes in the binding tendencies of multisensory signals. For instance, at the cognitive level one may ask whether an increased common source prior might result in an increased attention on the visual modality (and vica versa). Although, behavioural studies suggested that the ventriloquist effect is immune to spatial attention (Bertelson, Vroomen, et al., 2000; Vroomen et al., 2001b), no such studies exist for modality specific attention. Behavioural and neuroimaging studies have shown that selective attention can suppress the task-irrelevant sensory modality (Johnson & Zatorre, 2005; Laurienti et al., 2002; Mozolic et al., 2008; Santangelo & Macaluso, 2012). Also, it is well-established that

attentional resources can be adjusted flexibly to changes in environmental statistics for effective interactions with the environment (Dayan, Kakade, & Montague, 2000).

Similarly, in a dynamic multisensory environment the brain should allocate more attentional resources to the more spatially-reliable visual signal, if it is likely that the auditory and visual signals originate from the same source. Indeed, recent neuroimaging studies provide accumulating evidence that attention modulates audio-visual integration (Alsius et al., 2005; T. S. Andersen et al., 2004; Busse et al., 2005; Fairhall & Macaluso, 2009; Johnson & Zatorre, 2005; Nardo et al., 2014; Santangelo et al., 2009), including one study which examined the role of attention specifically in a ventriloquist situation (Busse et al., 2005). In a dynamically changing environment, we would speculate that allocation of greater attentional resources to the visual modality may increase the influence of the spatially more reliable visual signal on the perceived auditory location when the common source prior is high.

In conclusion, we demonstrate that human observers adapt their audio-visual binding tendencies dynamically to changes in the environmental statistics. In blocks where the probability of spatially collocated and synchronous audio-visual signals was high, observers bound audio-visual signals more often resulting in a higher ventriloquist illusion. Future studies will need to determine whether the changes in binding tendencies can be attributed to a dynamic allocation of attentional resources across the sensory modalities.

CHAPTER 5: VISUALLY INDUCED AUDITORY SPACE ADAPTATION

The current chapter is based on a project in collaboration with Mate Aller. The experiments were designed in a joint work. The pilot experiments were performed by the author. The implementation of the psychophysics and fMRI study was done by the author. All the analyses demonstrated in the thesis were performed by the author. The data of the psychophysics experiment were acquired by Mate Aller. The data of the fMRI experiment were acquired together. An additional EEG experiment (not demonstrated in the thesis) including the same participants as of those of the fMRI experiment is yet to be performed by Mate Aller.

Introduction

In order to make a robust percept of the environment, the brain integrates signals across the senses (Ernst & Bülthoff, 2004). The inherently noisy nature of sensory processing implies that sensory conflicts are always perceived in the nervous system. The temporary conflicts can be easily resolved across time scales; on the contrary, prolonged inter-sensory conflicts truly challenge the representations of the brain. A natural example of persistent audio-visual conflict occurs during development, when the brain needs to compensate for changes in eye separation and inter-aural differences (de Gelder & Bertelson, 2003). Developmental studies highlighted to the prominent role of recalibration before the age of 8 (Gori et al., 2008, 2010), and indeed a vast body of research demonstrated in the last fifty years that plasticity and adaptation is the key mechanism to resolve long-term inter-sensory conflicts (Ernst & Di Luca, 2011; Ghahramani et al., 1997; Held, 1965; Knudsen & Knudsen, 1989).

The spatial ventriloquist aftereffect is a prominent example of multisensory adaptation (Canon, 1970, 1971; Lewald, Foltys, & Töpper, 2002; Radeau & Bertelson, 1974, 1976; Recanzone, 1998). In a response to the prolonged exposure of spatial audio-visual conflict, the bran recalibrates the auditory space and the perceived locations of unisensory auditory signals will be shifted towards the previously displaced visual signal. Visually induced auditory recalibration has been suggested to occur rapidly (Bertelson, 1993; Lewald, 2002; Recanzone, 1998), yet, recalibration studies typically used adaptation periods of 10-30 min until recent years (Canon, 1970, 1971; Frissen et al., 2003; Lewald et al., 2002; Radeau & Bertelson, 1974, 1976; Recanzone, 1998; Woods & Recanzone, 2004). Lately, a series of studies demonstrated that auditory space aftereffects can be elicited after a very short adaptation period (Frissen et al., 2012; Mendonça et al., 2015; Wozny & Shams, 2011a), moreover, they indicated that a single audio-visual conflict is already sufficient to trigger recalibration (Wozny & Shams, 2011b). These new results corroborate very early findings almost 50 years ago describing that recalibration occurs as part of perceptual learning and as such it can be very rapid (Epstein, 1975). Wallach and colleagues were the first to demonstrate that binocular disparity and perceived depth can be modified by pairing the stimuli and recalibration effects were observed already after treatment of 2 min (Wallach, Moore, & Davidson, 1963).

Despite the numerous studies characterizing the ventriloquist aftereffect at the behavioural level, the neural substrates remain unknown. Studies have shown that auditory recalibration generalizes across space (Bertelson et al., 2006; Frissen et al., 2012). Findings are much more debated about the nature of generalization across auditory frequencies. Early studies (Lewald, 2002; Recanzone, 1998) demonstrated that recalibration does not generalize across frequencies, and based on this it was speculated that auditory adaptation occurs already

in the primary auditory cortex. The early stage processing was further supported by a recent EEG study, showing changes in event related potentials attributed to recalibration100 ms after auditory onset (Bruns, Liebnau, & Röder, 2011). On the contrary, multiple behavioural studies presented opposite findings revealing to the transfer of the ventriloquist aftereffect across wide frequency ranges (Frissen et al., 2003, 2005). In the lack of evidence about the brain structures involved in auditory adaptation, many candidate brain regions have been proposed over the years including the primary auditory cortex, planum temporale, parietal cortex and superior colliculus.

The present study combined psychophysics, fMRI and advanced multivariate decoding models to investigate visually induced auditory adaptation in the human brain focusing on the primary auditory cortex, planum temporale and intraparietal sulcus. Participants were presented with unisensory auditory signals before and after adaptation periods of spatially conflicting audio-visual signals. We demonstrated recalibration effects in a relatively large population in a psychophysics experiment, and selected a few participants to determine the neural substrates in a follow-up fMRI experiment. We trained support vector regression models on BOLD response patterns elicited by unisensory auditory stimuli before adaptation. Then we applied the learnt mapping to decode auditory spatial locations after adaptation. Fitting psychometric functions to participants' responded locations and neurometric functions to the BOLD-decoded locations we were able to characterize the remapping of auditory space both at the behavioural and neural level.

Experiment 1: Behavioural experiment

Methods

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 separate session before the main experiment to determine their performance in sound localization and adaptation tasks (for a detailed description, see the corresponding section). The study was approved by the human research ethics committee at the University of Birmingham.

Experimental procedure

Participants were screened in a separate session before the main experiment in sound localization as well as adaptation tasks of the main experiment. Participants selected for the experiment performed 4 sessions: 2 sessions with the V stimulus on the left during adaptation (VA adaptation with V stimulus on the left) and 2 sessions with the V stimulus on the right during adaptation (AV adaptation). The direction of adaptation was the same within a session to avoid any interaction between different directions of adaptation. The session order of adaptation was counter-balanced across participants.

Each session consisted of 5 pre-test periods followed by 10 adaptation periods interleaved by 10 post-test periods. A typical session lasted for ~1.5 hours and was divided up to 3 parts with breaks between. Each part consisted of 5 test periods (with an adaptation

period preceding each post-test period), and was completed in one sitting. The forehead of participants was marked at the beginning of the sessions to ensure the same head positioning throughout the session and avoid any variability in sound localization within a session due to head positioning. Participants were instructed to fixate to a central fixation cross throughout the session.

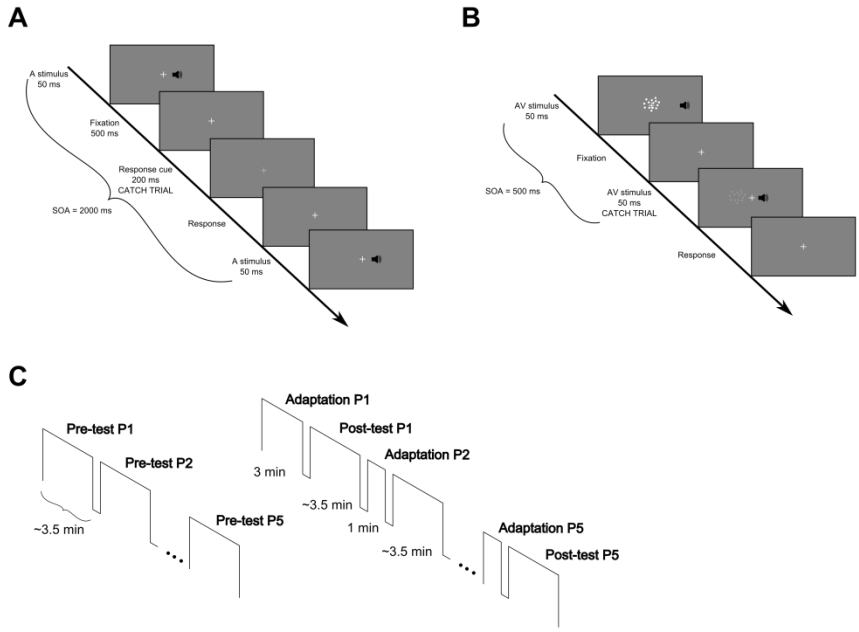


Figure 5.1 Experimental stimuli, tasks and design of the psychophysics experiment. (**A)** Example stimuli of the sound localization task. The first stimulus is a catch trial indicated by a 500 ms post-stimulus contrast change in the fixation cross. (**B)** Example stimuli of the adaptation task. The second AV stimulus is a catch trial indicated by a contrast change in the V stimulus. (**C)** Time course of the 3 phases: pre-test, adaptation and post-test. **Left:** Part 1 of each session consisting of 5 pre-test periods. **Right:** Parts 2 and 3 of each session consisting of 5-5 adaptation period interleaved by 5-5 post-test periods. Abbreviations: A, auditory; AV, audio-visual; SOA, stimulus onset asynchrony; P, period.

Stimuli

The visual (V) stimuli, used in the adaptation task, were cloud of 15 white dots (diameter=0.4° visual angle) sampled from a bivariate Gaussian presented on a dark grey background (90% contrast). 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 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 to the desired spatial locations. Scanner background noise was superimposed on the sound stimuli to match the task environment for the follow-up fMRI study.

Screening session to determine performance in sound localization and adaptation tasks

To familiarize participants with the sound localization task, they performed a 1-min practice run. Overall, participants performed 4 tasks during the screening session: (1) sound localization task of A stimuli responding on all trials; (2) sound localization task of A stimuli responding on catch trials; (3) sound localization task of audio-visual stimuli responding on all trials; and (4) adaptation task responding on catch trials. Task (2) and (4) were identical to the tasks of the main experiment. Task (1) was similar to task (2) with the exceptions of participants responding on all trials and the presentation of stimuli being not 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'>3.5 in task (2),
- d'>3.5 in task (4).

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

Task (4) served as a supplementary measure about participants' sound localization ability. Here, participants were presented with synchronous audio-visual stimuli for 50 ms with SOA=2.5 s. Audio-visual stimuli were sampled independently from 4 spatial locations (-9°, -3°, 3°, 9°) using 30 repetitions/condition amounting to 480 trials. Other aspects of the stimuli were identical to the audio-visual stimuli used in the main experiment.

Main experimental design

The experiment was divided up to 3 phases: pre-adaptation test, adaptation and post-adaptation test. The pre- and post-adaptation phases are called shortly pre- and post-tests or test phase, collectively. In the test phase, participants performed a sound localization task on A stimuli indicating a response only on catch trials. In the adaptation phase they performed a visual detection task on audio-visual stimuli.

In the sound localization task, participants were presented with A stimuli for 50 ms with SOA=1.8-2.2 s (using 0.4s uniform jitter). The stimuli were presented from ±12°, ±5°, ±2° and 0° visual angle. To match the task with the task of the follow-up fMRI study, stimuli were presented in blocks (~28 sec) separated by fixation periods (6 sec). In addition, stimuli were pseudo-randomized 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 (divided up to 5 stimulus blocks and 4 fixation periods) and lasted for ~3.5 min. Participants were instructed to localize the A stimuli, but indicating a response in a forced choice left-right discrimination task only on catch trials (20 out of 90 stimuli, ~22 %). Catch trials consisted of a 500 ms post-stimulus cue, a 55% contrast change in the fixation cross lasting for 200 ms. Participants were asked to use different hands in the 3 parts of the session.

Participants' hand responses were counter-balanced across sessions to avoid any motor response confound.

In the adaptation phase, participants were presented with audio-visual stimuli for 50 ms with SOA=0.5 s (without jitter). The first adaptation period (after break) consisted of 360 audio-visual stimuli (3 min), the other adaptations periods consisted of 120 audio-visual stimuli (1 min). The V stimuli were presented in three spatial locations (-5°, 0°, 5°) with 15° spatially discrepancy to the left (for VA adaptation) or right (for AV adaptation) of the A stimuli. Adaptation trials were 5 presentations of audio-visual stimuli from the same location, after which the stimulus location changed randomly. To ensure that participants attended to the stimuli there were occasional catch trials (10%), which consisted of a 20% contrast change of the V stimuli.

Eye movement recording and analysis

To address potential concerns that our 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 desktop mount of the Eyelink II system. Eye position data were automatically on-line parsed into events (saccade, fixation, eye blink) using the cognitive configuration of saccade detection (velocity threshold = 30°/sec, acceleration threshold = 8000°/sec^2, motion threshold = 0.15°). Fixation position was post-hoc offset corrected. Saccades were post-hoc filtered for radial amplitude larger than 1°. Eye data were analysed in the 0-500 msec post-stimulus period. Fixation was well maintained throughout the experiment with post-stimulus saccades detected in only 0.4% ± 0.1% (mean ± SEM) of all the trials. For the test phases, eye movement indices of % saccades, % eye blinks, and post-stimulus mean horizontal eye position were quantified and entered into one-way repeated measures ANOVAs. No significant differences were observed

across the conditions for % saccades and % blinks. A significant effect was found on the horizontal mean eye position ($F(2,14)=11.73$, $p=0.004$). Fisher's LSD post-hoc tests revealed differences between the pre-test and both post-test phases ($p=0.004$ and $p=0.027$ for post-test after VA and AV adaptation, respectively). We assume the finding is due to fatigue (looser fixation during post-tests) and should not be a real confound in our behavioural results.

Experimental setup

Visual and auditory stimuli were presented using Psychtoolbox version 3.0.11 (Brainard, 1997; Kleiner et al., 2007; Pelli, 1997) running under MATLAB R2011b (MathWorks Inc.) on a MacBook Pro (Mac OSX 10.6.8). Participants were seated at a distance of 60 cm from a 24'' LCD screen 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). Auditory stimuli were delivered via Sennheiser HD 280 Pro headphones. Participants' eye movements were recorded using an Eyelink II system (SR Research Ltd.) on a desktop mount at a sampling rate of 2000 Hz.

Analysis

Hit rates and signal sensitivity measure d-prime was calculated on catch trials for both sound localization and adaptation tasks as follows: $d' = Z(\text{probability}_{\text{hits}}) - Z(\text{probability}_{\text{false alarms}})$. 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).

In the sound localization task, cumulative Gaussian (*psychometric function*, PF) was fitted to the percentage 'perceived right responses' as a function of stimulus location (±12°, ±5°, ±2° and 0°) using maximum-likelihood estimation (www.palamedestoolbox.org). The

fitting procedure was based on fixed guess and lapse rates (each set to 0.02) and a constrained slope parameter resulting in estimates of the point of subjective equality (*PSE*), and the slope of the psychometric function. The slope estimate was converted to the just noticeable difference (*JND*) that is equivalent to the standard deviation of the psychometric function when defined at 84% level. For the screening session, PF was fitted on each task separately, on data pooled over periods. For the main experiment, PF was fitted separately for the pre- and post-test phases as well as each session, again, on data pooled over periods. The sessions of pre-tests (or post-tests) were fitted together by multi-condition model fitting using a constrained slope parameter. Fitting multiple models at the same time enabled to obtain less noisy PSE estimates. Also, fitting pre- and post-test data separately enabled to model possible temporal effects, given that pre-tests always preceded post-tests. Temporal effects could result either in improvement (e.g. due to training) or deterioration (e.g. due to fatigue) on task performance.

The aftereffect was calculated for each session by subtracting the PSE value of the pre-test phase from that of the post-test phase. This method enabled a more sensitive estimation of aftereffects not being masked by session to session variability in pre-test PSE values. Mean rightward and leftward aftereffects were calculated after AV and VA adaptations, respectively, averaging the values of the 2-2 sessions, respectively.

Additional analysis was also performed on aftereffects as a function of time. For this approach, the aftereffect was calculated for each post-test period subtracting the session-specific pre-test PSE value from those of the post-test periods. To account for the very noisy nature of PF fitting on single post-test periods (~3 responses/location), a fixed slope parameter was used taking the slope estimates of the overall post-test fits. We note that the

overall post-test fits were calculated using a constrained slope (see above), thus the slope estimate was one general value for each participant.

Results

In the current behavioural study, we employed a standard experimental design consisting of pre-test, adaptation and post-test phases to assess visually induced auditory adaptation. During pre- and post-tests, participants were presented with A stimuli from 7 horizontal spatial locations (±12°, ±5°, ±2° and 0°) and responded on catch trials in a forced-choice left-right discrimination task. Psychometric function (PF) was fitted to the percentage 'perceived right responses' as a function of stimulus location. During the adaptation phase, participants were presented with spatially discrepant audio-visual stimuli and attended the V stimuli performing a detection task. We expected that after adaptation participants perceive the A stimuli towards the displaced V stimuli indicating an aftereffect (AE) and resulting in the shift of the PF. Adaptation with V stimulus to the left of the sound (VA adaptation) would lead to a rightward shift of the function (due to fewer right responses), whilst adaptation with V stimulus on the right of the sound (AV adaptation) would lead to leftward shift of the function (due to more right responses).

In general, participants' performance was high on catch trials (>90%) detecting contrast changes of the visual stimuli and the fixation cross during the adaptation and the test phases, respectively (Table 5.1).

	Hit rate	**d-prime**	**PSE**
Pre-test	93.4% (±1.1%)	4.09 (±0.21)	0.58 (±0.21)
Adaptation			
AV adaptation	95.8% (±1.4%)	4.52 (±0.20)	not applicable

VA adaptation	95.6% (±1.6%)	4.62 (±0.22)	not applicable
Post-test			
After AV adaptation	91.5% (±1.7%)	4.02 (±0.20)	-0.99 (±0.29)
After VA adaptation	91.3% (±1.3%)	4.01 (±0.19)	2.11 (±0.25)

Table 5.1 Group-level mean values in hit rate, d-prime and PSE in the sound localization and adaptation tasks of the psychophysics experiment (SEM in parentheses).

Figue 5.2 displays the shifted psychometric functions of the participants. We can see that aftereffects were found both after AV adaptation (Figure 4.1A) and VA adaptation (Figure 4.1B) for each participant. This result is rather convincing, given that other studies reported occasional null or small negative effects with 1-2 subjects at this population size (Frissen et al., 2003; Mendonça et al., 2015). Group-level PSE values of the three test phases (pre-test, post-test after AV and VA adaptation) are summarized in Table 4.1. A paired t-test on the mean rightward and leftward aftereffects confirmed a highly significant result ($t(28)=11.512$, $p<0.0001$).

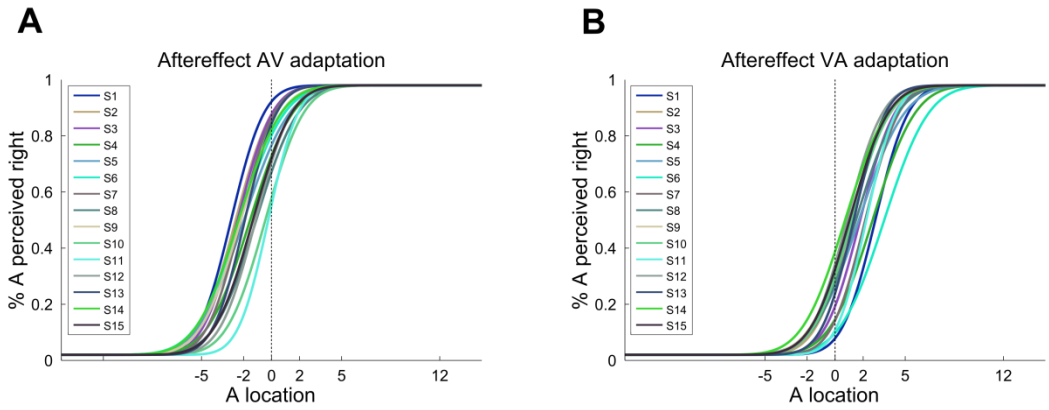


Figure 5.2 Psychometric function shifts after adaptation of the psychophysics experiment. (A) Psychometric functions are based on the mean aftereffects and the

constrained slope of each participant after AV adaptation. **(B)** Psychometric functions are based on the mean aftereffects and the constrained slope of each participant after VA adaptation. Abbreviations: A, auditory; AV, audiovisual; S, subject.

In addition, aftereffects were also evaluated in a time course analysis. Aftereffects were calculated for each post-test period to see whether: (i) the aftereffects were more prominent after 3 min adaptations (always in the beginning of parts in a session); and (ii) the interleaved 1 min adaptations are sufficient to maintain the aftereffect or there is any dissipation over time. The aftereffects of each period were entered into a 4 (sessions) x 2 (parts) x 5 (periods) repeated measures ANOVA. The main effect of session was highly significant ($F(3,14)=36.48$, $p<0.0001$) corresponding to our main finding showed earlier by paired t-tests. Apart from a weakly significant part x session interaction ($F(3,14)=3.15$, $p=0.037$), none of the other main effects or interactions were significant. These results suggest that the aftereffect was stable and did not dissipate over time. However, a null effect between the effectiveness of 1 and 3 min adaptations does not imply they are equally effective due to the exclusive precedence of 3 min adaptations before 1 min adaptations. In accordance, adaptation studies in recent years use a longer adaptation period for inducing an aftereffect and interleaved or shorter ones to maintain it (Bruns et al., 2011; Wozny & Shams, 2011a; Zaidel et al., 2013). Moreover, Frissen and colleagues demonstrated that 3 min adaptations induce larger aftereffects than 1 min adaptations (Frissen et al., 2012).

Experiment 2: fMRI experiment

Methods

Participants

Six right-handed participants (4 females, mean age=22.2; SD=3.7) with no history of neurological or psychiatric illness, and who showed the largest recalibration effects (for details, see Experiment 1 of the current chapter), were selected from the preceding behavioural experiment. One participant was excluded after 3 sessions due to an overall d'<2.5 in the adaptation tasks of the 3 sessions (same criterion was used for screening participants in the preceding 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.

Experimental procedure

Participants performed 4 sessions similarly to the preceding behavioural experiment. Here, the session order of adaptation was counter-balanced also within participants. Each session consisted of 10 pre-test periods followed by 8 adaptation periods interleaved by 8 post-test periods. In addition, periods were organized into runs and the scanner was restarted between each run. One run was composed of 2 pre-test periods *or* 2 adaptation periods and 2 post-test periods. Participants performed a short left-right discrimination task in the beginning of each session to check sufficient localization abilities in the scanner environment. Participants were instructed to fixate on a central fixation cross throughout the session.

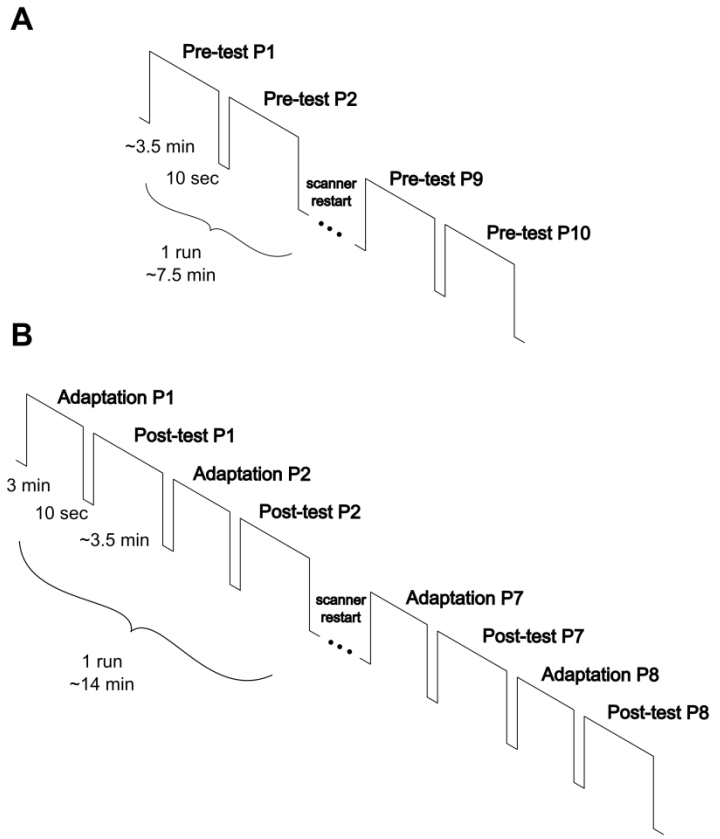


Figure 5.3 Design and time course of the fMRI experiment. (**A)** Pre-test phase of each session. (**B)** Adaptation and post-test phase of each session. Abbreviations: P, period.

Stimuli

Stimuli used in the fMRI experiment were identical to the preceding 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 preceding behavioural experiment except for 2 details of the adaptation task:

- All adaptation periods lasted for 3 min (except for the first 2 participants, where some 1 min adaptations were also used). This change enabled a more robust adaptation effect.
- Adaptation trials were 20 presentations of audio-visual stimuli from the same location, after which the audio-visual stimulus location changed randomly. This change was made to improve design efficiency for the audio-visual signals.

10 sec fixation was 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 preceding experiment except for the following: The visual stimuli were back projected to a Plexiglas screen using a D-ILA projector (JVC DLA-SX21) visible to the subject through a mirror mounted on the magnetic resonance (MR) head coil. Auditory stimuli were delivered via 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.

Behavioural analysis

The behavioural results obtained in the scanner were analysed similar as in the preceding behavioural study with the exception that no time course analysis was performed here.

MRI data acquisition

A 3T Philips Achieva scanner was used 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 mm^3 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 mm^3 voxels). There were 20 pre-test runs, each with 160 volumes over 4 sessions. There were 16 post-test runs both for left- and right-adaptations, over 2 sessions, respectively. The first 4 volumes were not acquired to allow T1 equilibration effects.

fMRI analysis

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, Holmes, Worsley, et al., 1995). 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 pre- and post-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. Condition-specific effects for each subject were estimated according to the general linear model (GLM).

The regressors of the sound location conditions (of the pre- and post-test phases) were used for multi-variate decoding analysis. We trained a linear support vector regression (SVR) model 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 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) from all but one pre-test run (leave-one-run out cross-validation, LORO CV). The model then used this learnt mapping to decode the spatial locations from the voxel response patterns of both the remaining pre-test run and all the post-test runs. The training-test procedure was repeated for all runs in a cross-validation scheme. This method yielded one estimated spatial location for each example of the pre-test runs and multiple estimates (as many CV folds) for examples of post-test runs. Default SVR hyper-parameters (C=0, ν=0.5) were used to train the models.

Inference was made based on the decoded spatial locations using *neurometric functions*. More specifically, the decoded spatial locations pooled over all CV folds were binarized as 'decoded right' and plotted as a function of sound locations. Similarly to the behavioural analysis, cumulative Gaussian was fitted to the percentage decoded right estimates as a function of stimulus location using maximum-likelihood estimation (www.palamedestoolbox.org). This method enabled a similar approach to the behavioural analysis, but now at the neural level. The fitting procedure was based on fixed guess and lapse rates (each set to 0.1) resulting in estimates of the point of subjective equality (*PSE*), and the slope of the psychometric function. The obtained neurometric functions (and their PSE values) of the participants were passed to the second level to allow for random effects analysis and inferences at the population level (Friston et al., 1999). Group-level mean and SEM neurometric functions were calculated from the pre-test, and two post-test (after left and right adaptation) neurometric functions of participants, and the corresponding PSE values were entered into a repeated measures ANOVA.

Region of interests used for decoding analysis

The auditory regions of interests (ROI) were defined based on using the Destrieux atlas of Freesurfer 5.3.0 (Dale et al., 1999; Destrieux et al., 2010). The regions were combined from the left and right hemispheres. The primary auditory cortex was based on the anterior transverse temporal gyrus (Heschl's gyrus). The higher auditory cortex was defined by merging the transverse temporal sulcus and the planum temporale. The intraparietal sulcus (IPS) was defined by merging the superior parietal gyrus and the intraparietal and sulcus labels. A sub-analysis in the preceding fMRI study (see Chapter 2) showed that decoding performance in the above mentioned anatomically defined IPS were reasonably similar to those obtained in the functionally defined IPS based on standard retinotopic mapping.

Results

In the fMRI study, we employed the same experimental design as in the preceding behavioural study consisting of pre-test, adaptation and post-test phases. The behavioural results in the scanner were evaluated similarly to before, fitting *psychometric functions* to the percentage 'perceived right responses' of participants as a function of stimulus location. We built support vector regression (SVR) models on the voxel response patterns in particular brain regions involved in auditory spatial processing. These models enabled to fit *neurometric functions* to the percentage 'decoded right choices' made by the linear support vector machine. We expected that neurometric functions would show similar shifts as the behavioural psychometric functions resulting in a rightward shift after adaptation with V stimulus to the left of the sound (VA adaptation) and a leftward shift after adaptation with V stimulus on the right of the sound (AV adaptation).

Behavioural results

Participants' performance in the scanner was high on catch trials (>89%) detecting contrast changes of the visual stimuli and the fixation cross during the adaptation and the test phases, respectively (Table 5.2).

	Hit rate	**d-prime**	**PSE**
Pre-test	97.5% (±0.5%)	4.8 (±0.16)	-0.73 (±1.09)
Adaptation			
Right-adaptation	92.4% (±0.7%)	4.54 (±0.05)	not relevant
Left-adaptation	89.5% (±1.6%)	4.09 (±0.13)	not relevant
Post-test			
After right-adaptation	97.8% (±0.4%)	5.01 (±0.13)	-2.54 (±0.52)
After left-adaptation	98.1% (±0.4%)	5.18 (±0.17)	2.46 (±0.16)

Table 5.2 Group-level mean values in hit rate, d-prime and PSE in the sound localization and adaptation tasks of the fMRI experiment (SEM in parentheses).

Figure 5.4 A and B displays the shifted psychometric functions after adaptation. As we expected, the functions and their corresponding PSE values moved consistently leftwards after AV adaptation due to the higher % right responses (Figure 5.4A). In contrast, the functions and the corresponding PSE values moved consistently rightwards after VA adaptation due to the lower % right responses (Figure 5.4B). Strong behavioural aftereffects were present in all sessions of the participants. Group-level PSE values of the three test phases (pre-test, post-test after AV and VA adaptation) are summarized in Table 5.2. A paired

t-test on the mean rightward and leftward aftereffects confirmed a highly significant result ($t(8)=9.187$, $p< 0.0001$).

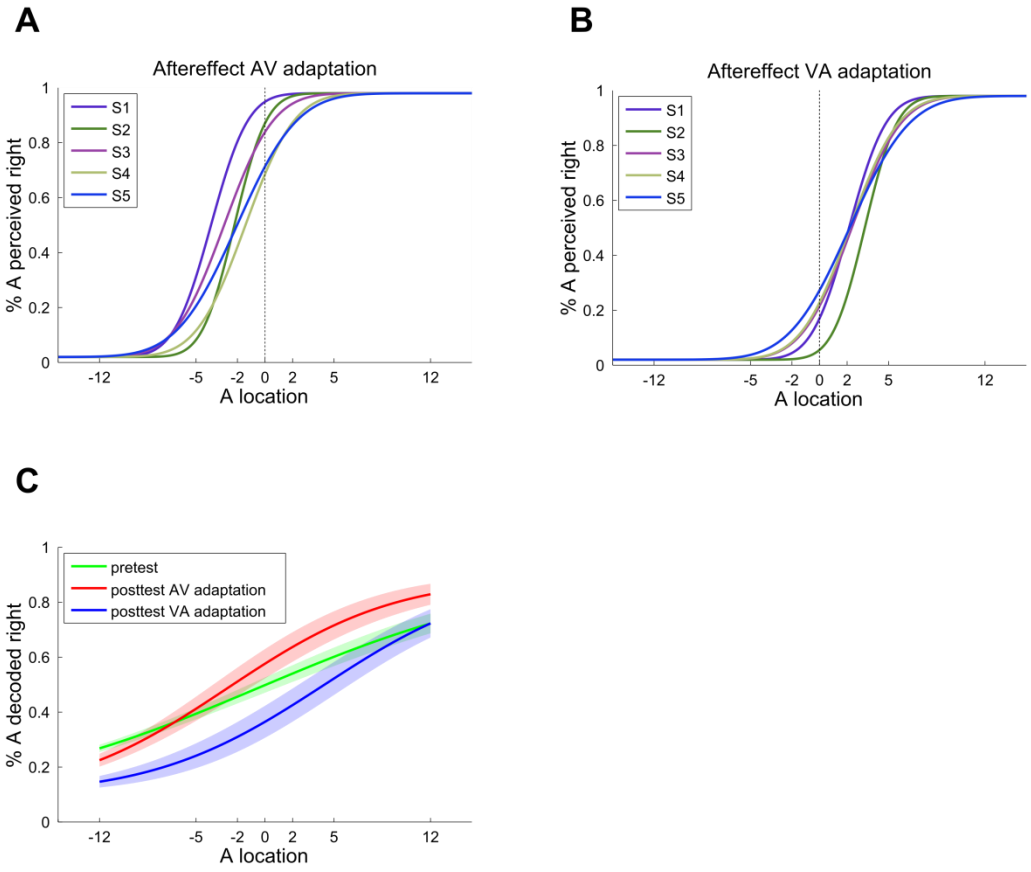


Figure 5.4 Psychometric and neurometric functions of the fMRI experiment. (A) Psychometric functions are based on the mean aftereffects and the constrained slope of each participant after AV adaptation. **(B)** Psychometric functions are based on the mean aftereffects and the constrained slope of each participant after VA adaptation. **(C)** Group-mean (±SEM) neurometric functions in planum temporale based on constrained post-test slope at the participant level. Abbreviations: A, auditory; AV, audiovisual; S, subject.

fMRI analysis: multivariate results

Decoding of auditory space: SVR models were trained on the pre-test auditory locations to find a mapping between BOLD response patterns and the stimulus locations (±12°, ±5°, ±2°

and 0°) using regressors based on responded and non-response trials. The SVR models were trained separately on our a priori regions of interest: Heschl's gyrus (HG; also primary auditory cortex, A1), planum temporale (PT) and intraparietal sulcus (IPS). The slope estimates of the participants' neurometric functions were passed to a between-subject 2nd level analysis. One sample t-tests revealed significant effects for all regions: Heschl's gyrus ($t(4) = 2.785$, $p<0.05$), planum temporale ($t(4) = 10.424$, $p<0.001$) and intraparietal sulcus ($t(4) = 5.160$, $p=0.007$). 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). We got similar results based on regressors of non-response trials.

Decoding of aftereffects: Using the SVR models built on the pre-test data, we decoded auditory stimulus locations after AV and VA adaptations. Only regressors from the non-response trials were used here to avoid any motor confound in the aftereffects. Again, we tested A1, PT and IPS separately. Aftereffects based on PSE estimates of participants' neurometric functions were passed to a between-subject 2nd level analysis. Paired t-tests revealed significant effect for planum temporale ($t(8) =4.978$, $p=0.002$; Figure 4.4C). A non-significant trend for PSE was also found in IPS (shifts in 4 out of 5 participants; $p>0.1$), and no effect was revealed in the primary auditory cortex.

Discussion

In order to adapt to the changes in our sensory systems or those in the environment, the brain needs to recalibrate the senses to keep them coordinated (Burr & Gori, 2012; de Gelder & Bertelson, 2003; Ernst & Di Luca, 2011). A vast body of behavioural research investigated spatial auditory adaptation induced by exposure to spatial audio-visual conflict, the so called ventriloquist aftereffect (Bertelson et al., 2006; Canon, 1970, 1971; Chen & Vroomen, 2013; Frissen et al., 2003, 2005; Lewald, 2002; Radeau & Bertelson, 1974, 1977; Recanzone, 1998;

Wozny & Shams, 2011b). Despite the overwhelming evidence accumulated by the behavioural studies, the neural structures remain unknown. The current study was designed to determine at which level of the auditory spatial processing visual signals induce adaptive changes in the neural representations. We used a classical recalibration paradigm with an adaptation period and test phases of sound localization before and after adaptation. In order to dissociate changes in neural representation of auditory space from changes due to motor responses, participants responded only on ~20% of trials in the sound localization task. These responses were used to confirm behavioural effects of our recalibration procedure. On the other side, trials where participants did not indicate a response inside the scanner were used to characterize the neural mechanisms of auditory spatial adaptation unconfounded by motor responses.

In our psychophysics experiment, we found very consistent recalibration effects in a population of fifteen healthy adults. In particular, when comparing the point of subjective equality of the psychometric functions after and before adaptation for each direction of adaptation, every single subject showed a recalibration effect. This consistency across participants confirms both the robustness of inter-sensory recalibration and the effectiveness of our experimental design. These results enabled to perform a fine tuned fMRI experiment to investigate the neural underpinnings of auditory space adaption on a five selected participants. In our fMRI experiment, at first we established whether auditory space can be decoded from our three candidate region of interests: primary auditory cortex (A1), planum temporale (PT) and intraparietal sulcus (IPS). These regions of interests were suggested both by behavioural studies investigating recalibration effects (Bruns & Röder, 2015; Frissen et al., 2005; Lewald, 2002; Recanzone, 1998) as well as they are based on neuroimaging studies characterizing auditory spatial processing in humans (Arnott, Binns, Grady, & Alain, 2004; Bushara et al.,

1999; Kong et al., 2014; Lewald et al., 2002; Michalka, Rosen, Kong, Shinn-Cunningham, & Somers, 2016; Weeks et al., 1999). To better understand these cortical regions we give an introduction to the dual-stream hypothesis. Originally, the dual-stream hypothesis was described for the visual system dividing it into two processing streams: a ventral non-spatial processing ('what') stream, and a dorsal spatial processing ('where') stream (Ungerleider & Haxby, 1994; Ungerleider & Mishkin, 1982). It has been proposed for auditory processing in non-human primates in the late 1990s (Kaas & Hackett, 2000; Rauschecker, 1998; Rauschecker & Tian, 2000; L M Romanski et al., 1999), and was soon extended to humans. The existence of the dual-stream hypothesis in humans was corroborated by a meta-analysis conducted on a total of 36 PET and fMRI studies in 2004 (Arnott et al., 2004). The ventral pathway is described to originate from the antero-lateral belt of the auditory cortex, propagates to the anterior part of the superior gyrus and terminates in the ventrolateral prefrontal (VLPFC). By contrast, the dorsal pathway originates from the dorso-lateral belt and parabelt (PT in humans), propagates to the posterior parietal cortex, and terminates in the dorsolateral prefrontal cortex (DLPFC). Over the years, some criticism has been raised regarding the specificity of these pathways, still the role of auditory spatial processing in the dorsal stream is generally accepted (Recanzone & Cohen, 2010 for a review). In recent years, the two-stream model has been extended to incorporate speech and music processing (Rauschecker & Scott, 2009), the role of the dorsal pathway in orienting vision has been emphasised (Arnott & Alain, 2011), and the parietal cortex has been demonstrated to be involved in supramodal spatial representations (Kong et al., 2014; Macaluso, Driver, & Frith, 2003; Michalka et al., 2016).

In line with previous research implicating that planum temporale has a prominent role in representing auditory space (Baumgart, Gaschler-Markefski, Woldorff, Heinze, & Scheich,

1999; Callan et al., 2015; Derey, Valente, de Gelder, & Formisano, 2016; Krumbholz, Schönwiesner, Rübsamen, et al., 2005; Krumbholz, Schönwiesner, Von Cramon, et al., 2005; Ortiz-Rios et al., 2017) the decoding accuracy for auditory locations was highest in planum temporale. Interestingly, we were able to decode auditory location significantly better than chance from the primary auditory cortex (Heschl's gyrus). This is supported by a clear evidence from a very recent neuropsychological study in macaque showing that hemifield code of the auditory space extends to the primary auditory cortices (Ortiz-Rios et al., 2017). Finally, we showed that auditory space could also be decoded from parietal cortex that is recognized as key a region for integrating spatial audio-visual signals into coherent representations. In summary, we were able to decode auditory spatial representations from all of three tested regions of interest (i.e. A1, PT, IPS) using multivariate models of BOLD-response patterns. Therefore, all of these regions may be also candidates for encoding visually induced changes in auditory spatial representations.

To characterize auditory space adaptation at the neural level we trained a support vector regression model for each subject on the auditory localization trials prior to adaptation to establish a mapping from BOLD-response patterns to spatial locations of the auditory space. Then using this learnt mapping we decoded BOLD-response patterns of the auditory signals after a visually induced spatial adaptation to characterize changes in auditory representation at the neural level. The neurometric functions (cumulative Gaussians fitted to the decoded locations) revealed a significant shift in point of subjective equality for left vs. right recalibration in PT consistently across all five participants. At a lesser extent, but recalibration induced a consistent change in PSE values in IPS in four out of five subjects. Neurometric functions in the primary auditory cortex were not significantly different after left and right visual adaptation. 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 PT and multisensory association areas as IPS.

Future investigations need to determine the neural mechanisms underlying audio-visual recalibration and address how these adaptive changes in auditory representation emerge over time. For instance, IPS as a multisensory convergence zone might play a key role in comparing visual and auditory signals and induce neuroplasticity via top-down effects on PT. Alternatively, plasticity may be induced directly in PT and then propagate to IPS. We also need to examine carefully the role of other cortical (e.g. FEF) and subcortical regions (e.g. superior colliculus), whether they contribute to the adaptation processes.

CHAPTER 6: GENERAL DISCUSSION AND CONCLUSIONS

The work presented in this thesis aimed to understand how the brain adapts to changes in the multisensory environment. In particular, we investigated the neural mechanisms of audio-visual integration (Chapter 3) and changes in the neural representation of auditory space due to permanent spatial conflict between the audio-visual signals (Chapter 5). In a behavioural study, we investigated how the human brain adapts to temporary changes in the statistics of audio-visual signals (Chapter 4). In this final chapter, I will summarize the main findings of the empirical chapters, discuss how they contribute to the literature and outline directions for future research.

Overview of findings

Chapter 3: Neural basis of explicit causal inference in audio-visual perception

The human brain is in a state of constant adaptation: it has to accommodate changes in the sensory systems as well as those in the environment. To form a veridical representation of the external world, the brain needs to infer the causal structure of the world needs and signals should be integrated only if they belong to the same object otherwise kept separate (Trommershäuser et al., 2011). Although numerous behavioural studies investigated such a causal judgement, the neural substrates of explicit causal inference on audio-visual signals remain unknown. I investigated the neural representations of spatial auditory and visual signals, and how their interaction results as well as is influenced by explicit causal judgements of the signals. Crucially, our design allowed us to dissociate causal judgements both from the physical congruency of the audio-visual signals and the selected motor actions.

We demonstrated that the dorsolateral prefrontal cortex (DLPFC) is the key region in perceptual judgement (i.e. causal inference judgement). We found that the spatial

representation of auditory and visual signals as well as their interaction (i.e. physical congruency) are encoded at multiple levels of the cortical hierarchy from low-level sensory cortices to the parietal (IPS) and prefrontal (FEF, DLPFC) regions. Interestingly, AV signals interacted already in planum temporale, which is considered to be a low-level sensory cortex for auditory processing. This finding is consistent with several anatomical, neuroimaging and electrophysiological studies showing cross-modal interplay at early sensory cortices (Driver & Noesselt, 2008; Ghazanfar & Schroeder, 2006).

In line with our finding of AV interactions at multiple levels of the cortex, recent neuroimaging evidence showed that spatial representations are hierarchically organized in the human brain (Rohe & Noppeney, 2015a, 2016). AV spatial representations start with unisensory estimates in the low level cortical areas, in IPS0-2 bisensory representations are based on weighting the reliabilities of the signals, finally IPS3-4 forms spatial estimates also taking into account the implicit causal relationship of the signals (Rohe & Noppeney, 2015a). Importantly, it seems plausible that the evidence about the world's causal structure accumulated in DLPFC is backwards projected to these regions to inform their spatial representations. Indeed, we were able to decode perceptual congruency not only from DLPFC but V2-3 and planum temporale suggesting these top-down modulations.

Furthermore, we demonstrated that FEF and IPS form a circuitry where all aspects of spatial representation (visual, auditory, motor, physical and perceptual congruency) are encoded. This finding suggests two roles for these regions: (i) they serve as convergence zones for bottom-up sensory correspondences and top-down influences; and (ii) they also transform audio-visual spatial representations into motor responses according to arbitrary mappings. These findings are consistent with the role of parietal cortex forming priority maps and guiding motor action (Bisley & Goldberg, 2010; Cohen, 2009; Sereno & Huang, 2014).

Chapter 4: Changing the tendency to integrate audio-visual signals

Having characterized the neural mechanisms of explicit causal inference, we asked whether the causal judgement, i.e. perception of common source audio-visual signals is stable or susceptible to changes. In particular, we were interested if changes in the statistics of audio-visual stimuli (e.g. frequency of signals coming from the same source) can change the expectations of the brain, which in turn would lead to changes in the binding tendencies.

In a behavioural study, we manipulated the relative frequencies of trials where audio-visual signals were collocated and synchronous ('common source signals') or spatially disparate and asynchronous ('separate source signals') in short blocks of 32 trials. We demonstrated that observers bind audio-visual signals more often in blocks with high frequency of spatiotemporally congruent trials as indexed by the ventriloquist effect. Critically, given the short blocks these adaptive changes were rapid and also without much of awareness of the observers.

These results converge with a previous finding from Van Wanrooij and colleagues (Van Wanrooij et al., 2010) showing reduced response times when the probability of congruent trials was increased in a ventriloquist situation. Also, similar results have been shown using the McGurk illusion, when a series of synchronous audio-visual movies resulted in an increased binding of the signals (Gau & Noppeney, 2016; Nahorna et al., 2012, 2015).

Collectively, these results suggest that the dynamic update of common source prior is a generic mechanism for audio-visual integration. Importantly, changing the tendency to integrate audio-visual signals can be considered as an adaptive process to sudden changes in the environmental statistics.

Chapter 5: Visually induced auditory space adaptation

What happens when persistent changes occur in the environmental statistics, e.g. sensory conflict is maintained over time? The brain adaptively recalibrates the senses to keep them coordinated (Burr & Gori, 2012; de Gelder & Bertelson, 2003; Ernst & Di Luca, 2011). We studied the ventriloquist aftereffect as the most prominent audio-visual adaptation paradigm (Canon, 1970, 1971; Lewald et al., 2002; Radeau & Bertelson, 1974, 1976; Recanzone, 1998). We aimed to characterize the brain regions underlying of the adaptation processes that in spite of the overwhelming behavioural evidence remain unknown.

First, we demonstrated that the auditory space can be decoded from all of our a priori tested cortical regions, such as the primary auditory cortex, planum temporale and intraparietal sulcus. The role of planum temporale in encoding auditory space is in line with a vast body of research (Baumgart et al., 1999; Callan et al., 2015; Derey et al., 2016; Krumbholz, Schönwiesner, Rübsamen, et al., 2005; Krumbholz, Schönwiesner, Von Cramon, et al., 2005; Ortiz-Rios et al., 2017), whilst the contribution of IPS is more sparse in the literature (Lewald et al., 2002; Lewald, Riederer, Lentz, & Meister, 2008). Similarly, the contribution of primary auditory cortex is not well established, however, there has been a very recent clear evidence in monkey (Ortiz-Rios et al., 2017).

We used the above mentioned auditory mappings from stimulus location to BOLD activation patterns to examine which of these region(s) are involved in the remapping of auditory space after adaptation. We demonstrated clear shifts in the neurometric functions of planum temporale in each fMRI participant indicating adaptation effects in this region. Similar results were observed in IPS in four out of five participants, although the effect was not significant due to the small sample size. Crucially, all these results manifest auditory space maps and their adaptation in a manner that is not confounded by motor responses.

Contributions, and future directions

The contributions of this thesis are addressed from multiple points of view. In each view, I will discuss the relevant connections between the empirical chapters and put them into the context of the literature. I will point out agreements and discrepancies; and possibly I will address intriguing directions for future research.

In *Chapter 3*, we investigated the neural mechanisms of explicit causal inference based on bottom-up correspondences of audio-visual signals, such as spatial disparity. The results suggested three key players in causal inference, DLPFC, IPS and FEF. I proposed that at the top of the cortical hierarchy, DLPFC infers the causal structure of audio-visual signals and top-down modulates both the FEF-IPS circuitry and other lower level sensory regions (e.g. planum temporale). The FEF-IPS circuitry forms spatial representations both based on bottom-up sensory inputs and these top-down causal inferences. These results extend recent neuroimaging findings showing the role of IPS3-4 in forming AV spatial representations taking into account the implicit causal structure of the signals (Rohe & Noppeney, 2015a). In *Chapter 4*, we took a different approach and investigated whether causal inference ('common source judgement') can be driven by top-down mechanisms built up from expectations on recent audio-visual experiences. Indeed, we demonstrated that observers dynamically adapted their binding tendency ('common source prior') based on frequency changes of spatiotemporally congruent and incongruent trials. Similarly, numerous studies presented evidence that human learn and update priors dynamically (Adams et al., 2004; Berniker et al., 2010; Knill, 2007; Körding, Ku, & Wolpert, 2004; Körding & Wolpert, 2004; Sato & Körding, 2014). Our results are also in line with previous audio-visual research showing increased occurrence of McGurk illusion in the context of synchronous audio-visual movies (Gau & Noppeney, 2016; Nahorna et al., 2012, 2015). On the contrary, a recent study

demonstrated that the observer's tendency to bind audio-visual signals is stable across days (Odegaard & Shams, 2016). The discrepancy can be resolved taking into account the different time scales and the environmental statistics of audio-visual signals. On one hand, a stable binding tendency over longer periods of time and in the same environment enables a robust and reliable integration. On the other hand, the flexibility to adjust binding tendencies to sudden changes in statistical regularities of audio-visual signals (e.g. more frequent co-occurrence) allows optimal integration in different environments. Future studies will need to address the cognitive and neural mechanisms of binding tendencies.

The different time scales discussed just before also brings us to the next topic: the time scales of adaptation. As the results of *Chapter 4* showed and discussed above, updating binding tendencies allows a rapid adaptation to temporary changes in the statistics of audio-visual signals. Another type of adaptation occurs on longer time scales as revealed in *Chapter 5*: the brain recalibrates the senses to each other. Although, recent evidence suggests that recalibration can also occur rapidly (Frissen et al., 2012; Mendonça et al., 2015; Wozny & Shams, 2011b), the real benefit of recalibration becomes prominent on longer time scales e.g. during development (Burr et al., 2011). The long-lasting nature of sensory recalibration is supported by the maintained aftereffect in our experiments. Especially, in the psychophysics experiment, where initial 3 min adaptation periods followed by 3.5 min sound localization test phases and interleaved 1 min adaptations were sufficient to maintain aftereffects for ~20-25 min. The maintenance of recalibration by such a rapid and effective way explains its rising popularity in recent years that allowed the characterization of the underlying computational and some neuronal mechanisms (Bruns et al., 2011; Wozny & Shams, 2011a; Zaidel et al., 2013, 2011). Zaidel and colleagues demonstrated that unsupervised recalibration follows a fix ratio adaptation and argued that recalibration should be based on accuracy, not reliability

(Zaidel et al., 2011). In a follow-up study they investigated supervised calibration providing feedback about accuracy (Zaidel et al., 2013). Although, interesting questions arise e.g. are there ways that observers can estimate accuracy without external feedback? Can unreliable cues recalibrate other senses in an unsupervised manner? Are their findings indeed general as they suggested or specific to visual-vestibular recalibration?

Finally, let me discuss the neuronal representations of audio-visual space from an adaptive point of view. In *Chapter 3*, I demonstrated that a wide range of system of brain regions is sensitive to physical congruency, such as spatial disparity. This system includes mostly higher order association cortices as IPS, FEF and DLPFC. Putatively, this system is mainly focused on translating physical congruency into perceptual congruency and does not have strong spatial representations. On the other side, low level sensory cortices (e.g. V1-3, belt regions in auditory cortex, planum temporale) have strong spatial representations. Interestingly, planum temporale belongs to both groups. Therefore, one might speculate that it is an ideal candidate for adaptation, when its prominent representations get modulated by cross-modal (e.g. visual) interactions. Indeed, I demonstrated in *Chapter 5*, that planum temporale decodes auditory space more precisely than primary auditory cortex or intraparietal sulcus, and it is the region where recalibration effects were most prominent. Future research needs to investigate the exact mechanisms of these interactions, and test whether cognitive factors e.g. attention could potentially play a role in these processes

Conclusions

In summary, the current thesis aimed to characterize the neural processes underlying audio-visual integration and adaptation. I demonstrated that dorsolateral prefrontal cortex infers the explicit causal structure of audio-visual signals, by which they were generated. Two further key players in the human cortical hierarchy are the frontal eye fields and intraparietal sulcus,

where bottom-up audio-visual spatial information and possibly top-down modulated explicit causal inference converge to inform spatial representations and map them into motor actions. I showed that human observers dynamically adapt to changes in the environmental statistics of audio-visual signals. I proposed that they achieve it by updating their prior assumption on the causal structure (i.e. 'common source prior') of audio-visual signals. I observed mappings of auditory space at three levels of the auditory cortical hierarchy: in primary auditory cortex, planum temporale and intraparietal sulcus. Moreover, I described remapping at the behavioural and neural level followed by visually induced auditory space adaptation. Collectively, these results extend our understanding of short-term audio-visual adaptation and present neural mechanisms underlying audio-visual integration and adaptation.

REFERENCES

Adam, R., & Noppeney, U. (2014). A phonologically congruent sound boosts a visual target into perceptual awareness. *Frontiers in Integrative Neuroscience*, *8*(September), 70.

Adams, W. J., Graf, E. W., & Ernst, M. O. (2004). Experience can change the "light-from-above" prior. *Nature Neuroscience*, *7*(10), 1057–1058.

Alais, D., & Burr, D. (2004). The ventriloquist effect results from near-optimal bimodal integration. *Current Biology : CB*, *14*(3), 257–62.

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.

Allefeld, C., & Haynes, J.-D. (2014). Searchlight-based multi-voxel pattern analysis of fMRI by cross-validated MANOVA. *NeuroImage*, *89*, 345–57.

Alsius, A., Navarra, J., Campbell, R., & Soto-Faraco, S. (2005). Audiovisual integration of speech falters under high attention demands. *Current Biology*, *15*, 839–843.

Andersen, R. A., Snyder, L. H., Bradley, D. C., & Xing, J. (1997). Multimodal representation of space in the posterior parietal cortex and its use in planning movements. *Annual Review of Neuroscience*, *20*, 303–30.

Andersen, T. S., Tiippana, K., & Sams, M. (2004). Factors influencing audiovisual fission and fusion illusions. *Cognitive Brain Research*, *21*, 301–308.

Arnott, S. R., & Alain, C. (2011). The auditory dorsal pathway: Orienting vision. *Neuroscience and Biobehavioral Reviews*, *35*(10), 2162–2173.

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.

Ashburner, J., & Friston, K. J. (2005). Unified segmentation. *NeuroImage*, *26*(3), 839–851.

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.

Barnes, C. L., & Pandya, D. N. (1992). Efferent cortical connections of multimodal cortex of the superior temporal sulcus in the rhesus monkey. *The Journal of Comparative Neurology*, *318*(2), 222–44.

Barraclough, N. E., Xiao, D. K., 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.

Baumgart, F., Gaschler-Markefski, B., Woldorff, M. G., Heinze, H. J., & Scheich, H. (1999). A movement-sensitive area in auditory cortex. *Nature*, *400*(6746), 724–726.

Beauchamp, M. S., Lee, K., Argall, B., & Martin, A. (2004). Integration of auditory and visual information about objects in superior temporal sulcus. *Neuron*, *41*, 809–823.

Bedford, F. L. (2001). Towards a general law of numerical/object identity. *Cahiers de Psychologie Cognitive/Current Psychology of Cognition*, *20*, 113–176.

Beierholm, U. R., Quartz, S., & Shams, L. (2009). Bayesian priors are encoded independently from likelihoods in human multisensory perception. *Journal of Vision*, *9*, 23.1-9.

Bental, E., Dafny, N., & Feldman, S. (1968). Convergence of auditory and visual stimuli on single cells in the primary visual cortex of unanesthetized unrestrained cats. *Exp Neurol*, *20*(3), 341–351.

Berniker, M., Voss, M., & Körding, K. P. (2010). Learning priors for bayesian computations in the nervous system. *PLoS ONE*, *5*(9), 1–9.

Bernstein, L. E., Auer, E. T., & Takayanagi, S. (2004). Auditory speech detection in noise enhanced by lipreading. *Speech Communication*, *44*(1–4 SPEC. ISS.), 5–18.

Bertelson, P. (1993). The time-course of adaptation to auditory–visual spatial discrepancy. In *Proc. 6th Conf. Eur. Soc. Cognit. Psychol.* Copenhagen.

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

Bertelson, P., Frissen, I., Vroomen, J., & de Gelder, B. (2006). The aftereffects of ventriloquism: patterns of spatial generalization. *Perception & Psychophysics*, *68*(3), 428–436.

Bertelson, P., Pavani, F., Ladavas, E., Vroomen, J., & de Gelder, B. (2000). Ventriloquism in patients with unilateral visual neglect. *Neuropsychologia*, *38*(12), 1634–42.

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

Bertelson, P., & Radeau, M. (1981). Cross-modal bias and perceptual fusion with auditory-visual spatial discordance. *Perception & Psychophysics*, *29*(6), 578–84.

Bertelson, P., Vroomen, J., de Gelder, B., & Driver, J. (2000). The ventriloquist effect does not depend on the direction of deliberate visual attention. *Perception & Psychophysics*, *62*(2), 321–32.

Besle, J., Fort, A., Delpuech, C., & Giard, M. H. (2004). Bimodal speech: Early suppressive visual effects in human auditory cortex. *European Journal of Neuroscience*, *20*(8), 2225–2234.

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), 578–86.

Bisley, J. W., & Goldberg, M. E. (2010). Attention, intention, and priority in the parietal lobe. *Annual Review of Neuroscience*, *33*, 1–21.

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.

Blauert, J. (1969). Sound localization in the median plane. *Acta Acustica United with Acustica*, *22*(4), 205–213.

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*, 425–34.

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 : CB*, *17*(19), 1697–703.

Brainard, D. H. (1997). The Psychophysics Toolbox. *Spatial Vision*, *10*, 433–436.

Bremmer, F., Schlack, A., Shah, N. J., Zafiris, O., Kubischik, M., Hoffmann, K.-P., … Fink, G. R. (2001). Polymodal Motion Processing in Posterior Parietal and Premotor Cortex. *Neuron*, *29*, 287–296.

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), 554–64.

Brosch, M., Selezneva, E., & Scheich, H. (2005). Nonauditory events of a behavioral procedure activate auditory cortex of highly trained monkeys. *The Journal of Neuroscience : The Official Journal of the Society for Neuroscience*, *25*(29), 6797–806.

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.

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.

Bruns, P., Maiworm, M., & Röder, B. (2014). Reward expectation influences audiovisual spatial integration. *Attention, Perception & Psychophysics*, *76*(6), 1815–1827.

Bruns, P., & Röder, B. (2015). Sensory recalibration integrates information from the immediate and the cumulative past. *Scientific Reports*, *5*, 12739.

Budinger, E., Heil, P., Hess, A., & Scheich, H. (2006). Multisensory processing via early cortical stages: Connections of the primary auditory cortical field with other sensory systems. *Neuroscience*, *143*(4), 1065–1083.

Burr, D., Binda, P., & Gori, M. (2011). Multisensory Integration and Calibration in Adults and in Children. In *Sensory Cue Integration* (Vol. 6, pp. 173–194). Oxford University Press.

Burr, D., & Gori, M. (2012). Multisensory integration develops late in humans. In *The neural bases of multisensory processes.* (pp. 1–21).

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.

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 of the United States of America*, *102*(51), 18751–6.

Butler, R. A. (1969). Monaural and binaural localization of noise bursts vertically in the median sagittal plane. *J. Audit. Res.*, *3*, 230–235.

Butler, R. A. (1971). The monaural localization of tonal stimuli. *Percept. Psychophys.*, *9*(1B), 99–101.

Calhoun, V. D., & Adali, T. (2006). Unmixing fMRI with independent component analysis. *IEEE Engineering in Medicine and Biology Magazine : The Quarterly Magazine of the Engineering in Medicine & Biology Society*, *25*(2), 79–90.

Callan, A., Callan, D., & Ando, H. (2015). An fMRI Study of the Ventriloquism Effect. *Cerebral Cortex (New York, N.Y. : 1991)*, (Jeffress 1948), 1–11.

Calvert, G. A., Brammer, M. J., Bullmore, E., Campbell, R., Iversen, S. D., & David, a S. (1999). Response amplification in sensory-specific cortices during crossmodal binding. *Neuroreport*, *10*(12), 2619–2623.

Calvert, G. A., Bullmore, E., Brammer, M. J., Campbell, R., Williams, S. C., McGuire, P. K., … David, A. S. (1997). Activation of auditory cortex during silent lipreading. *Science*, *276*(5312), 593–596.

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.

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.

Canon, L. K. (1971). Directed attention and maladaptive “adaptation” to displacement of the visual field. *Journal of Experimental Psychology*, *88*(3), 403–408.

Chang, C.-C., & Lin, C.-J. (2011). LIBSVM: A library for support vector machines. *ACM Transactions on Intelligent Systems and Technology*, *2*(3), 1–27.

Chen, L., & Vroomen, J. (2013). Intersensory binding across space and time: a tutorial review. *Attention, Perception & Psychophysics*, *75*(5), 790–811.

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

Cohen, Y. E. (2009). Multimodal activity in the parietal cortex. *Hearing Research*, *258*(1–2), 100–105.

Cohen, Y. E., & Andersen, R. A. (2004). Multisensory representations of space in the posterior parietal cortex. In G. A. Calvert, C. Spence, & B. E. Stein (Eds.), *The Handbook of multisensory processes* (pp. 463–482). London: The MIT Press.

Colin, C., Radeau, M., Soquet, A., Dachy, B., & Deltenre, P. (2002). Electrophysiology of spatial scene analysis: The mismatch negativity (MMN) is sensitive to the ventriloquism illusion. *Clinical Neurophysiology*, *113*, 507–518.

Cox, D. D., & Savoy, R. L. (2003). Functional magnetic resonance imaging (fMRI) "brain reading": Detecting and classifying distributed patterns of fMRI activity in human visual cortex. *NeuroImage*, *19*(2), 261–270.

Dale, A. M., Fischl, B., & Sereno, M. I. (1999). Cortical Surface-Based Analysis: I. Segmentation and Surface Reconstruction. *NeuroImage*, *9*(2), 179–194.

Dayan, P., Kakade, S., & Montague, P. R. (2000). Learning and selective attention. *Nature Neuroscience*, *3*(Supp), 1218–1223.

de Gelder, B., & Bertelson, P. (2003). Multisensory integration, perception and ecological validity. *Trends in Cognitive Sciences*, *7*(10), 460–467.

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.

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.

Diedrichsen, J., & Kriegeskorte, N. (2017). Representational models: A common framework for understanding encoding, pattern-component, and representational-similarity analysis. *PLoS Computational Biology*, *13*(4), e1005508.

Dixon, W. J., & Mood, A. M. (1948). A method for obtaining and analyzing sensitivity data. *Journal of the American Statistical Association*, *43*(241), 109.

Dong, C., Swindale, N. V, & Cynader, M. S. (1999). A contingent aftereffect in the auditory system. *Nature Neuroscience*, *2*(10), 863–865.

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

Driver, J., & Noesselt, T. (2008). Multisensory interplay reveals crossmodal influences on "sensory-specific" brain regions, neural responses, and judgments. *Neuron*, *57*(1), 11–23.

Epstein, W. (1975). Recalibration by pairing: a process of perceptual learning. *Perception*, *4*(1), 59–72.

Eramudugolla, R., Kamke, M. R., Soto-Faraco, S., & Mattingley, J. B. (2011). Perceptual load influences auditory space perception in the ventriloquist aftereffect. *Cognition*, *118*(1), 65–77.

Ernst, M. O., & Banks, M. S. (2002). Humans integrate visual and haptic information in a statistically optimal fashion. *Nature*, *415*(6870), 429–33.

Ernst, M. O., & Bülthoff, H. H. (2004). Merging the senses into a robust percept. *Trends in Cognitive Sciences*, *8*(4), 162–9.

Ernst, M. O., & Di Luca, M. (2011). Multisensory Perception: From Integration to Remapping. In *Sensory Cue Integration* (pp. 224–250).

Fairhall, S. L., & Macaluso, E. (2009). Spatial attention can modulate audiovisual integration at multiple cortical and subcortical sites. *European Journal of Neuroscience*, *29*(December 2008), 1247–1257.

Falchier, A., Clavagnier, S., Barone, P., & Kennedy, H. (2002). Anatomical evidence of multimodal integration in primate striate cortex. *The Journal of Neuroscience : The Official Journal of the Society for Neuroscience*, *22*(13), 5749–59.

Fechner, G. T. (1860). *Elemente der Psychophysik*. Leipzig: Breitkopf und Härtel.

Feddersen, W. E., Sandel, T. T., Teas, D. C., & Jeffress, L. A. (1957). Localization of high-frequency tones. *The Journal of the Acoustical Society of America*, *29*(9), 988–991.

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

Fishman, M. C., & Michael, C. R. (1973). Integration of Auditory Information in the Cat's Visual Cortex. *Vision Research*, *13*(8), 1415–1419.

Frissen, I., de Gelder, B., & Vroomen, J. (2012). The Aftereffects of Ventriloquism: The Time Course of the Visual Recalibration of Auditory Localization. *Seeing and Perceiving*, *25*(1), 1–14.

Frissen, I., Vroomen, J., de Gelder, B., & Bertelson, P. (2003). The aftereffects of ventriloquism: are they sound-frequency specific? *Acta Psychologica*, *113*(3), 315–27.

Frissen, I., Vroomen, J., de Gelder, B., & Bertelson, P. (2005). The aftereffects of ventriloquism: generalization across sound-frequencies. *Acta Psychologica*, *118*(1–2), 93–100.

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.

Friston, K. J., Holmes, A. P., Price, C. J., Büchel, C., & Worsley, K. J. (1999). Multisubject fMRI studies and conjunction analyses. *NeuroImage*, *10*(4), 385–396.

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., Jezzard, P., & Turner, R. (1994). Analysis of functional MRI time-series. *Human Brain Mapping*, *1*(2), 153–171.

Friston, K. J., Worsley, K. J., Frackowiak, R. S. J., Mazziotta, J. C., & Evans, A. C. (1994). Assessing the significance of focal activations using their spatial extent. *Human Brain Mapping*, *1*(3), 210–220.

Fu, K.-M. G., Shah, A. S., O'Connell, M. N., McGinnis, T., Eckholdt, H., Lakatos, P., … Schroeder, C. E. (2004). Timing and laminar profile of eye-position effects on auditory responses in primate auditory cortex. *Journal of Neurophysiology*, *92*(6), 3522–3531.

Fuster, J. M., Bodner, M., & Kroger, J. K. (2000). Cross-modal and cross-temporal association in neurons of frontal cortex. *Nature*, *405*(6784), 347–51.

García-Pérez, M. A. (1998). Forced-choice staircases with fixed step sizes: asymptotic and small-sample properties. *Vision Research*, *38*(12), 1861–81.

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

Gau, R., & Noppeney, U. (2016). How prior expectations shape multisensory perception. *NeuroImage*, *124*, 876–886.

Gazzaniga, M. S. (2000). *The New Cognitive Neurosciences* (2nd edn). MIT press.

Ghahramani, Z., Wolpert, D. M., & Jordan, M. I. (1997). Computational models of sensorimotor integration. In *Advances in Psychology* (Vol. 119, pp. 117–147).

Ghazanfar, A. A., Maier, J. X., Hoffman, K., & Logothetis, N. K. (2005). Multisensory Integration of Dynamic Faces and Voices in Rhesus Monkey Auditory Cortex. *Journal of Neuroscience*, *25*(20), 5004–5012.

Ghazanfar, A. A., & Schroeder, C. E. (2006). Is neocortex essentially multisensory? *Trends in Cognitive Sciences*, *10*(6), 278–85.

Giard, M. H., & Peronnet, F. (1999). Auditory-visual integration during multimodal object recognition in humans: a behavioral and electrophysiological study. *Journal of Cognitive Neuroscience*, *11*(5), 473–90.

Gondan, M., Niederhaus, B., Rösler, F., & Röder, B. (2005). Multisensory processing in the redundant-target effect: A behavioral and event-related potential study. *Perception & Psychophysics*, *67*(4), 713–726.

Gori, M., Del Viva, M., Sandini, G., & Burr, D. (2008). Young Children Do Not Integrate Visual and Haptic Form Information. *Current Biology*, *18*(9), 694–698.

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.

Gottlieb, J., & Snyder, L. H. (2010). Spatial and non-spatial functions of the parietal cortex. *Current Opinion in Neurobiology*, *20*(6), 731–740.

Graziano, M. S. A., Reiss, L. A. J., & Gross, C. G. (1999). A neuronal representation of the location of nearby sounds. *Nature*, *397*(6718), 428–30.

Graziano, M. S. A., Yap, G. S., & Gross, C. G. (1994). Coding of visual space by premotor neurons. *Science (New York, N.Y.)*, *266*(5187), 1054–1057.

Hackett, T. A., Stepniewska, I., & Kaas, J. H. (1998). Subdivisions of auditory cortex and ipsilateral cortical connections of the parabelt auditory cortex in macaque monkeys. *The Journal of Comparative Neurology*, *394*(4), 475–95.

Hall, J. L. (1981). Hybrid adaptive procedure for estimation of psychometric functions. *The Journal of the Acoustical Society of America*, *69*(6), 1763–1769.

Hanke, M., Halchenko, Y. O., Sederberg, P. B., Hanson, S. J., Haxby, J. V, & Pollmann, S. (2009). PyMVPA: A python toolbox for multivariate pattern analysis of fMRI data. *Neuroinformatics*, *7*(1), 37–53.

Hansen, L. K., Larsen, J., Nielsen, F. A., Strother, S. C., Rostrup, E., Savoy, R., … Paulson, O. B. (1999). Generalizable patterns in neuroimaging: how many principal components? *Neuroimage*, *9*(5), 534–544.

Hanson, S. J., & Halchenko, Y. O. (2008). Brain reading using full brain support vector machines for object recognition: there is no “face” identification area. *Neural Computation*, *20*(2), 486–503.

Hastie, T., Friedman, J., & Tibshirani, R. (2001). *The elements of statistical learning.* New York, NY: Springer New York.

Haxby, J. V. (2001). Distributed and Overlapping Representations of Faces and Objects in Ventral Temporal Cortex. *Science*, *293*(5539), 2425–2430.

Haynes, J.-D. (2015). A Primer on Pattern-Based Approaches to fMRI: Principles, Pitfalls, and Perspectives. *Neuron*, *87*(2), 257–270.

Haynes, J.-D., & Rees, G. (2006). Decoding mental states from brain activity in humans. *Nature Reviews. Neuroscience*, *7*(7), 523–34.

Hays, W. L. (1994). *Statistics*. Wadsworth Publishing Company.

Hebart, M. N., Görgen, K., Haynes, J.-D., & Dubois, J. (2015). The Decoding Toolbox

(TDT): a versatile software package for multivariate analyses of functional imaging data. *Frontiers in Neuroinformatics*, *8*(January), 1–18.

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

Held, R. (1965). Plasticity in sensory-motor systems. *Scientific American*, *213*(5), 84–94.

Howard, I. P., & Templeton, W. B. (1966). *Human Spatial Orientation*. New York: Wiley.

Johnson, J. A., & Zatorre, R. J. (2005). Attention to simultaneous unrelated auditory and visual events: Behavioral and neural correlates. *Cerebral Cortex*, *15*(October), 1609–1620.

Kaas, J. H., & Hackett, T. A. (2000). Subdivisions of auditory cortex and processing streams in primates. *Proc Natl Acad Sci U S A*, *97*(22), 11793–11799.

Kaernbach, C. (1991). Simple adaptive testing with the weighted up-down method. *Perception & Psychophysics*, *49*(3), 227–9.

Kawaura, J., Suzuki, Y., Asano, F., & Sone, T. (1991). Sound localization in headphone reproduction by simulating transfer functions from the sound source to the external ear. *Journal of the Acoustical Society of Japan (E)*, *12*(5), 203–216.

Kayser, C., Petkov, C. I., Augath, M., & Logothetis, N. K. (2007). Functional Imaging Reveals Visual Modification of Specific Fields in Auditory Cortex. *J Neurosci*, *27*(1824--1835), 1824–1835.

Kayser, C., Petkov, C. I., Remedios, R., & Logothetis, N. K. (2012). Multisensory Influences on Auditory Processing: Perspectives from fMRI and Electrophysiology. In *The Neural Bases of Multisensory Processes* (pp. 1–12).

Kingdom, F. A. A., & Prins, N. (2010). *Psychophysics: a practical introduction*. Academic Press.

Kleiner, M., Brainard, D. H., & Pelli, D. G. (2007). What's new in Psychtoolbox-3?

Knill, D. C. (2007). Learning Bayesian priors for depth perception. *Journal of Vision*, *7*(8), 13.

Knudsen, E. I. (2002). Instructed learning in the auditory localization pathway of the barn owl. *Nature*, *417*(6886), 322–8.

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), 3306–13.

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.

Kontsevich, L. L., & Tyler, C. W. (1999). Bayesian adaptive estimation of psychometric slope and threshold. *Vision Research*, *39*(16), 2729–37.

Körding, K. P., Beierholm, U. R., Ma, W. J., Quartz, S., Tenenbaum, J. B., & Shams, L. (2007). Causal Inference in Multisensory Perception. *PLoS ONE*, *2*(9), 10.

Körding, K. P., Ku, S., & Wolpert, D. M. (2004). Bayesian integration in force estimation. *Journal of Neurophysiology*, *92*(5), 3161–5.

Körding, K. P., & Wolpert, D. M. (2004). Bayesian integration in sensorimotor learning. *Nature*, *427*(January), 244–247.

Kriegeskorte, N., Goebel, R., & Bandettini, P. A. (2006). Information-based functional brain mapping. *Proceedings of the National Academy of Sciences of the United States of America*, *103*(10), 3863–8.

Krugliak, A., & Noppeney, U. (2016). Synaesthetic interactions across vision and audition. *Neuropsychologia*, *88*, 65–73.

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.

Krumbholz, K., Schönwiesner, M., Von Cramon, D. Y., Rübsamen, R., Shah, N. J., Zilles, K., & Fink, G. R. (2005). Representation of interaural temporal information from left and right auditory space in the human planum temporale and inferior parietal lobe. *Cerebral Cortex*, *15*(3), 317–324.

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), 279–92.

Laurienti, P. J., Burdette, J. H., Wallace, M. T., Yen, Y.-F., Field, A. S., & Stein, B. E. (2002). Deactivation of sensory-specific cortex by cross-modal stimuli. *Journal of Cognitive Neuroscience*, *14*, 420–429.

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.

Lee, H., & Noppeney, U. (2011). Physical and perceptual factors shape the neural mechanisms that integrate audiovisual signals in speech comprehension. *The Journal of Neuroscience : The Official Journal of the Society for Neuroscience*, *31*(31), 11338–50.

Lee, H., & Noppeney, U. (2014). Music expertise shapes audiovisual temporal integration windows for speech, sinewave speech, and music. *Frontiers in Psychology*, *5*(AUG), 1–9.

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.

Leinonen, L., Hyvärinen, J., & Sovijärvi, A. R. A. (1980). Functional properties of neurons in the temporo-parietal association cortex of awake monkey. *Experimental Brain Research*, *39*(2), 203–215.

Lewald, J. (2002). Rapid adaptation to auditory-visual spatial disparity. *Learning & Memory (Cold Spring Harbor, N.Y.)*, *9*(5), 268–78.

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.

Lewald, J., Riederer, K. a J., Lentz, T., & Meister, I. G. (2008). Processing of sound location in human cortex. *The European Journal of Neuroscience*, *27*(5), 1261–70.

Lomo, T., & Mollica, A. (1959). [Activity of single units of the primary optic cortex during stimulation by light, sound, smell and pain, in unanesthetized rabbits]. *Bollettino Della Societa Italiana Di Biologia Sperimentale*, *35*, 1879–82.

Macaluso, E., Driver, J., & Frith, C. D. (2003). Multimodal spatial representations engaged in human parietal cortex during both saccadic and manual spatial orienting. *Current Biology : CB*, *13*(12), 990–9.

Magnotti, J. F., Ma, W. J., & Beauchamp, M. S. (2013). Causal inference of asynchronous audiovisual speech. *Frontiers in Psychology*, *4*(NOV), 1–10.

Maiworm, M., Bellantoni, M., Spence, C., & Röder, B. (2012). When emotional valence modulates audiovisual integration. *Attention, Perception, & Psychophysics*, *74*, 1302–1311.

Martuzzi, R., Murray, M. M., Michel, C. M., Thiran, J.-P., Maeder, P. P., Clarke, S., & Meuli, R. (2007). Multisensory interactions within human primary cortices revealed by BOLD

dynamics. *Cerebral Cortex*, *17*(7), 1672–1679.
McAlpine, D. (2005). Creating a sense of auditory space. *The Journal of Physiology*, *566*(Pt 1), 21–8.
McGurk, H., & Macdonald, J. (1976). Hearing lips and seeing voices. *Nature*, *264*, 691–811.
Mendonça, C. (2014). A review on auditory space adaptations to altered head-related cues. *Frontiers in Neuroscience*, *8*(8 JUL), 1–14.
Mendonça, C., Escher, A., van de Par, S., & Colonius, H. (2015). Predicting auditory space calibration from recent multisensory experience. *Experimental Brain Research*, 1983–1991.
Meredith, M. A., & Stein, B. E. (1983). Interactions among converging sensory inputs in the superior colliculus. *Science*, *221*(4608), 389–391.
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.
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.
Middlebrooks, J. C. (1999). Virtual localization improved by scaling nonindividualized external-ear transfer functions in frequency. *The Journal of the Acoustical Society of America*, *106*(3 Pt 1), 1493–510.
Middlebrooks, J. C., & Green, D. M. (1991). Sound localization by human listeners. *Annual Review of Psychology*, *42*(234046), 135–59.
Miller, L. M. (2005). Perceptual Fusion and Stimulus Coincidence in the Cross-Modal Integration of Speech. *Journal of Neuroscience*, *25*(25), 5884–5893.
Mitchell, T., Hutchinson, R., Niculescu, R. S., Pereira, F., Wang, X., Just, M., & Newman, S. (2004). Learning to decode cognitive states from brain images. *Machine Learning*, *57*(1/2), 145–175.
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.
Moller, H., Sorensen, M. F., Jensen, C. B., & Hammershoi, D. (1996). Binaural technique: do we need individual recordings? *Journal of the Audio Engineering Society*, *44*(6), 451–69.
Moore, B. C. J. (2013). *An Introduction to the Psychology of Hearing*. *The Journal of the Acoustical Society of America* (6th ed.). Boston: Brill.
Morrell, F. (1972). Visual system's view of acoustic space. *Nature*, *238*, 44–46.
Mozolic, J. L., Hugenschmidt, C. E., Peiffer, A. M., & Laurienti, P. J. (2008). Modality-specific selective attention attenuates multisensory integration. *Experimental Brain Research*, *184*, 39–52.
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.
Mur, M., Bandettini, P. A., & Kriegeskorte, N. (2009). Revealing representational content with pattern-information fMRI - An introductory guide. *Social Cognitive and Affective Neuroscience*, *4*(1), 101–109.
Murata, K., Cramer, H., & Bach-y-Rita, P. (1965). Neuronal convergence of noxious, acoustic, and visual stimuli in the visual cortex of the cat. *Journal of Neurophysiology*, *28*(6), 1223–39.
Musacchia, G., & Schroeder, C. E. (2009). Neuronal mechanisms, response dynamics and perceptual functions of multisensory interactions in auditory cortex. *Hearing Research*,

258(1–2), 72–9.
Nahorna, O., Berthommier, F., & Schwartz, J.-L. (2012). Binding and unbinding the auditory and visual streams in the McGurk effect. *The Journal of the Acoustical Society of America*, *132*(2), 1061–77.
Nahorna, O., Berthommier, F., & Schwartz, J.-L. (2015). Audio-visual speech scene analysis: Characterization of the dynamics of unbinding and rebinding the McGurk effect. *The Journal of the Acoustical Society of America*, *137*(May), 362–377.
Nardo, D., Santangelo, V., & Macaluso, E. (2014). Spatial orienting in complex audiovisual environments. *Human Brain Mapping*, *35*(4), 1597–1614.
Naselaris, T., Kay, K. N., Nishimoto, S., & Gallant, J. L. (2011). Encoding and decoding in fMRI. *NeuroImage*, *56*(2), 400–410.
Nichols, T. E., & Holmes, A. P. (2002). Nonparametric permutation tests for functional neuroimaging: a primer with examples. *Human Brain Mapping*, *15*(1), 1–25.
Noppeney, U., Josephs, O., Hocking, J., Price, C. J., & Friston, K. J. (2008). The effect of prior visual information on recognition of speech and sounds. *Cerebral Cortex*, *18*(3), 598–609.
Noppeney, U., Ostwald, D., & Werner, S. (2010). Perceptual decisions formed by accumulation of audiovisual evidence in prefrontal cortex. *The Journal of Neuroscience : The Official Journal of the Society for Neuroscience*, *30*(21), 7434–46.
Norman, K. a, Polyn, S. M., Detre, G. J., & Haxby, J. V. (2006). Beyond mind-reading: multi-voxel pattern analysis of fMRI data. *Trends in Cognitive Sciences*, *10*(9), 424–30.
Odegaard, B., & Shams, L. (2016). The Brains Tendency to Bind Audiovisual Signals Is Stable but Not General. *Psychological Science*.
Odegaard, B., Wozny, D. R., & Shams, L. (2015). Biases in Visual, Auditory, and Audiovisual Perception of Space. *PLoS Computational Biology*, *11*(12), 1–23.
Odegaard, B., Wozny, D. R., & Shams, L. (2016). The effects of selective and divided attention on sensory precision and integration. *Neuroscience Letters*, *614*, 24–28.
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*, 1–13.
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.
Pekkola, J., Ojanen, V., Autti, T., Jääskeläinen, I. P., Möttönen, R., Tarkiainen, A., & Sams, M. (2005). Primary auditory cortex activation by visual speech: an fMRI study at 3 T. *NeuroReport*, *16*(2), 125.
Pelli, D. G. (1997). The VideoToolbox software for visual psychophysics: transforming numbers into movies. *Spatial Vision*, *10*, 437–442.
Pereira, F., & Botvinick, M. (2011). Information mapping with pattern classifiers: a comparative study. *NeuroImage*, *56*(2), 476–96.
Pereira, F., Mitchell, T., & Botvinick, M. (2009). Machine learning classifiers and fMRI: a tutorial overview. *NeuroImage*, *45*(1 Suppl), S199-209.
Radeau, M., & Bertelson, P. (1974). The after-effects of ventriloquism. *The Quarterly Journal of Experimental Psychology*, *26*(1), 63–71.
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.
Radeau, M., & Bertelson, P. (1977). Adaptation to auditory-visual discordance and ventriloquism in semirealistic situations. *Perception & Psychophysics*, *22*(2), 137–146.

Radeau, M., & Bertelson, P. (1978). Cognitive factors and adaptation to auditory-visual discordance. *Perception & Psychophysics*, *23*(4), 341–3.

Radeau, M., & Bertelson, P. (1987). Auditory-visual interaction and the timing of inputs. Thomas (1941) revisited. *Psychological Research*, *49*(1), 17–22.

Rajkowska, G., & Goldman-Rakic, P. S. (1995). Cytoarchitectonic definition of prefrontal areas in the normal human cortex: II. Variability in locations of areas 9 and 46 and relationship to the Talairach Coordinate System. *Cerebral Cortex (New York, N.Y. : 1991)*, *5*(4), 323–37.

Rauschecker, J. P. (1998). Parallel processing in the auditory cortex of primates. *Audiol Neurootol*, *3*(2–3), 86–103.

Rauschecker, J. P., & Scott, S. K. (2009). Maps and streams in the auditory cortex: Nonhuman primates illuminate human speech processing. *Nature Neuroscience*, *12*(6), 718–24.

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.

Rayleigh, L. (1907). On our perception of sound direction. *Philosophical Magazine*, *13*, 214–232.

Recanzone, G. H. (1998). Rapidly induced auditory plasticity: the ventriloquism aftereffect. *Proceedings of the National Academy of Sciences of the United States of America*, *95*(3), 869–75.

Recanzone, G. H. (2009). Interactions of auditory and visual stimuli in space and time. *Hearing Research*, *258*(1–2), 89–99.

Recanzone, G. H., & Cohen, Y. E. (2010). Serial and parallel processing in the primate auditory cortex revisited. *Behavioural Brain Research*, *206*(1), 1–7.

Recanzone, G. H., & Sutter, M. L. (2008). The biological basis of audition. *Annual Review of Psychology*, *59*, 119–42.

Repp, B. H. (1984). Categorical perception: Issues, methods, findings. *Speech and Language: Advances in Basic Research and Practice*.

Roach, N. W., Heron, J., & McGraw, P. V. (2006). Resolving multisensory conflict: a strategy for balancing the costs and benefits of audio-visual integration. *Proceedings. Biological Sciences / The Royal Society*, *273*(1598), 2159–2168.

Rock, I., & Victor, J. (1964). Vision and touch: an experimentally created conflict between the two senses. *Science (New York, N.Y.)*, *143*(3606), 594–6.

Rockland, K. S., & Ojima, H. (2003). Multisensory convergence in calcarine visual areas in macaque monkey. *International Journal of Psychophysiology*, *50*(1–2), 19–26.

Rohe, T., & Noppeney, U. (2015a). Cortical Hierarchies Perform Bayesian Causal Inference in Multisensory Perception. *PLOS Biology*, *13*, e1002073.

Rohe, T., & Noppeney, U. (2015b). Sensory reliability shapes Bayesian Causal Inference in perception via two mechanisms. *Journal of Vision*, *15*(2015), 1–38.

Rohe, T., & Noppeney, U. (2016). Distinct computational principles govern multisensory integration in primary sensory and association cortices. *Current Biology*, *26*(4), 509–514.

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.

Romanski, L. M., Tian, B., Fritz, J., Mishkin, M., Goldman-Rakic, P. S., & Rauschecker, J. P. (1999). Dual streams of auditory afferents target multiple domains in the primate prefrontal cortex. *Nature Neuroscience*, *2*(12), 1131–6.

Rowland, B., Stanford, T. R., & Stein, B. E. (2007). A Bayesian model unifies multisensory

spatial localization with the physiological properties of the superior colliculus. *Experimental Brain Research*, *180*, 153–161.
Russo, G. S., & Bruce, C. J. (1994). Frontal eye field activity preceding aurally guided saccades. *Journal of Neurophysiology*, *71*(3), 1250–1253.
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).
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., Olivetti Belardinelli, M., Spence, C., & Macaluso, E. (2009). Interactions between voluntary and stimulus-driven spatial attention mechanisms across sensory modalities. *Journal of Cognitive Neuroscience*, *21*, 2384–2397.
Sato, Y., & Körding, K. P. (2014). How much to trust the senses: Likelihood learning. *Journal of Vision*, *14*(13), 1–13.
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), 3335–55.
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.
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), 319–37.
Schwartz, J.-L., Berthommier, F., & Savariaux, C. (2004). Seeing to hear better: Evidence for early audio-visual interactions in speech identification. *Cognition*, *93*(2), 69–78.
Sereno, M. I., Dale, A. M., Reppas, J. B., Kwong, K. K., Belliveau, J. W., Brady, T. J., … Tootell, R. B. (1995). Borders of multiple visual areas in humans revealed by functional magnetic resonance imaging. *Science (New York, N.Y.)*, *268*(5212), 889–93.
Sereno, M. I., & Huang, R.-S. (2014). Multisensory maps in parietal cortex. *Current Opinion in Neurobiology*, *24*(1), 39–46.
Shams, L., & Beierholm, U. R. (2010). Causal inference in perception. *Trends in Cognitive Sciences*, *14*(9), 425–32.
Shams, L., Ma, W. J., & Beierholm, U. R. (2005). Sound-induced flash illusion as an optimal percept. *Neuroreport*, *16*(17), 1923–7.
Slutsky, D. A., & Recanzone, G. H. (2001). Temporal and spatial dependency of the ventriloquism effect. *Neuroreport*, *12*(1), 7–10.
Smiley, J., Hackett, T. A., Ulbert, I., Karmos, G., Lakatos, P., Javitt, D. C., & Schroeder, C. E. (2007). Multisensory convergence in auditory cortex, I. Cortical connections of the caudal superior temporal plane in macaque monkeys. *The Journal of Comparative Neurology*, *502*(6), 894–923.
Spence, C. (2009). *Explaining the Colavita visual dominance effect*. *Progress in Brain Research* (Vol. 176). Elsevier.
Spence, C., & Driver, J. (2000). Attracting attention to the illusory location of a sound: reflexive crossmodal orienting and ventriloquism. *Neuroreport*, *11*(9), 2057–2061.
Spinelli, D. N. (1968). Auditory specificity in uniy recordings from cat's visual cortex. *ExpNeurology*, *22*, 75–84.
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.
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.
Stein, B. E., & Meredith, M. A. (1993). *The merging of the senses.* Cambridge MA: The MIT Press.
Stekelenburg, J. J., Vroomen, J., & de Gelder, B. (2004). Illusory sound shifts induced by the ventriloquist illusion evoke the mismatch negativity. *Neuroscience Letters*, *357*, 163–166.
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.
Stevenson, R. A., Geoghegan, M. L., & James, T. W. (2007). Superadditive BOLD activation in superior temporal sulcus with threshold non-speech objects. *Experimental Brain Research*, *179*(1), 85–95.
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. *The Journal of Neuroscience : The Official Journal of the Society for Neuroscience*, *26*(43), 11138–47.
Swisher, J. D., Halko, M. A., Merabet, L. B., McMains, S. A., & Somers, D. C. (2007). Visual topography of human intraparietal sulcus. *The Journal of Neuroscience : The Official Journal of the Society for Neuroscience*, *27*(20), 5326–37.
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.
Teder-Sälejärvi, W. A., Russo, F. Di, McDonald, J. J., Hillyard, S. A., & Di Russo, F. (2005). Effects of Spatial Congruity on Audio-Visual Multimodal Integration. *Journal of Cognitive Neuroscience*, *17*(9), 1396–1409.
Thomas, G. J. (1941). Experimental study of the influence of vision on sound localization. *Journal of Experimental Psychology*, *28*(2), 163–177.
Tong, F., & Pratte, M. S. (2012). Decoding patterns of human brain activity. *Annual Review of Psychology*, *63*, 483–509.
Trommershäuser, J., Kording, K., & Landy, M. S. (Eds.). (2011). *Sensory Cue Integration*. Oxford University Press.
Ungerleider, L., & Haxby, J. V. (1994). "What" and "where" in the human brain. *Current Opinion in Neurobiology*, *4*(2), 157–165.
Ungerleider, L., & Mishkin, M. (1982). Two cortical visual systems. *Analysis of Visual Behavior*.
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.
Van Wanrooij, M. M., Bremen, P., & John Van Opstal, A. (2010). Acquired prior knowledge modulates audiovisual integration. *The European Journal of Neuroscience*, *31*(10), 1763–71.
van Wassenhove, V. (2013). Speech through ears and eyes: Interfacing the senses with the supramodal brain. *Frontiers in Psychology*, *4*(JUL), 1–17.
van Wassenhove, V., Grant, K. W., & Poeppel, D. (2005). Visual speech speeds up the neural processing of auditory speech. *Proceedings of the National Academy of Sciences of the United States of America*, *102*(4), 1181–6.
von Saldern, S., & Noppeney, U. (2013). Sensory and striatal areas integrate auditory and visual signals into behavioral benefits during motion discrimination. *The Journal of Neuroscience : The Official Journal of the Society for Neuroscience*, *33*(20), 8841–9.

Vroomen, J., Bertelson, P., & de Gelder, B. (2001a). Directing spatial attention towards the illusory location of a ventriloquized sound. *Acta Psychologica*, *108*(1), 21–33.

Vroomen, J., Bertelson, P., & de Gelder, B. (2001b). The ventriloquist effect does not depend on the direction of automatic visual attention. *Perception & Psychophysics*, *63*(4), 651–9.

Vroomen, J., & de Gelder, B. (2004). Perceptual Effects of Cross-Modal Stimulation: Ventriloquism the Freezing Phenomenon. In *The handbook of multisensory processes* (pp. 141–150).

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. Experimentelle Hirnforschung. Expérimentation Cérébrale*, *158*(2), 252–8.

Wallach, H., Moore, M. E., & Davidson, L. (1963). Modification of stereoscopic depth-perception. *The American Journal of Psychology*, *76*(2), 191–204.

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.

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.

Warren, D. H., Welch, R. B., & McCarthy, T. J. (1981). The role of visual-auditory "compellingness" in the ventriloquism effect: implications for transitivity among the spatial senses. *Perception & Psychophysics*, *30*(6), 557–564.

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.

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

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.

Welch, R. B. (1999). Meaning, attention, and the "unity assumption" in the intersensory bias of spatial and temporal perceptions. In *Cognitive contributions to the perception of spatial and temporal events* (pp. 371–387). Amsterdam: North-Holland/Elsevier Science Publishers.

Welch, R. B., & Warren, D. H. (1980). Immediate perceptual response to intersensory discrepancy. *Psychological Bulletin*, *88*(3), 638–67.

Wenzel, E. M., Arruda, M., Kistler, D. J., & Wightman, F. L. (1993). Localization using nonindividualized head-related transfer functions. *The Journal of the Acoustical Society of America*, *94*(1), 111–23.

Werner, S., & Noppeney, U. (2010). Superadditive responses in superior temporal sulcus predict audiovisual benefits in object categorization. *Cerebral Cortex (New York, N.Y. : 1991)*, *20*(8), 1829–42.

Werner, S., & Noppeney, U. (2011). The contributions of transient and sustained response codes to audiovisual integration. *Cerebral Cortex (New York, N.Y. : 1991)*, *21*(4), 920–31.

Wetherill, G. B., & Levitt, H. (1965). Sequential estimation of points on a psychometric function. *The British Journal of Mathematical and Statistical Psychology*, *18*(1), 1–10.

Wichmann, F. A., & Hill, N. J. (2001a). The psychometric function: I. Fitting, sampling, and goodness of fit. *Perception & Psychophysics*, *63*(8), 1293–313.

Wichmann, F. A., & Hill, N. J. (2001b). The psychometric function: II. Bootstrap-based

confidence intervals and sampling. *Perception & Psychophysics*, *63*(8), 1314–29.
Wightman, F. L., & Kistler, D. J. (1997). Monaural sound localization revisited. *The Journal of the Acoustical Society of America*, *101*(2), 1050–1063.
Witten, I. B., & Knudsen, E. I. (2005). Why seeing is believing: Merging auditory and visual worlds. *Neuron*, *48*(3), 489–496.
Woods, T. M., & Recanzone, G. H. (2004). Visually induced plasticity of auditory spatial perception in macaques. *Current Biology : CB*, *14*(17), 1559–64.
Worsley, K. J., & Friston, K. J. (1995). Analysis of fMRI time-series revisited--again. *NeuroImage*, *2*(3), 173–81.
Wozny, D. R., Beierholm, U. R., & Shams, L. (2008). Human trimodal perception follows optimal statistical inference. *Journal of Vision*, *8*(3), 24.1-11.
Wozny, D. R., Beierholm, U. R., & Shams, L. (2010). Probability matching as a computational strategy used in perception. *PLoS Computational Biology*, *6*(8).
Wozny, D. R., & Shams, L. (2011a). Computational Characterization of Visually Induced Auditory Spatial Adaptation. *Frontiers in Integrative Neuroscience*, *5*(November), 1–11.
Wozny, D. R., & Shams, L. (2011b). Recalibration of auditory space following milliseconds of cross-modal discrepancy. *The Journal of Neuroscience : The Official Journal of the Society for Neuroscience*, *31*(12), 4607–12.
Yendiki, A., Greve, D. N., Wallace, S., Vangel, M., Bockholt, J., Mueller, B. A., … Gollub, R. L. (2010). Multi-site characterization of an fMRI working memory paradigm: reliability of activation indices. *NeuroImage*, *53*(1), 119–31.
Yuille, A., & Bülthoff, H. H. (1996). Bayesian Decision Theory and Psychophysics. In D. C. Knill & W. Richards (Eds.), *Bayesian Approaches to Perception* (pp. 123–163). Cambridge: Cambridge University Press.
Zaidel, A., Ma, W. J., & Angelaki, D. E. (2013). Supervised calibration relies on the multisensory percept. *Neuron*, *80*(6), 1544–57.
Zaidel, A., Turner, A. H., & Angelaki, D. E. (2011). Multisensory calibration is independent of cue reliability. *The Journal of Neuroscience : The Official Journal of the Society for Neuroscience*, *31*(39), 13949–62.