Multisensory Processing in the Ferret Auditory Cortex



Amy J Hammond-Kenny

Merton College, University of Oxford

A thesis submitted for the degree of

Doctor of Philosophy

Trinity 2017

Multisensory Processing in the Ferret Auditory Cortex

Amy J Hammond-Kenny

Merton College, University of Oxford

A thesis submitted for the degree of Doctor of Philosophy

Trinity, 2017

Our perception of events depends on the integration of information derived from different sensory modalities. Functional imaging and electrophysiological studies have shown that multisensory interactions occur even at the level of primary sensory cortices; however their functional significance remains elusive. Therefore, to explore the relationship between multisensory interactions in early auditory cortical processing areas and behaviour, we recorded activity from the auditory cortex during the performance of different auditory-visual tasks.

Ferrets were trained by positive operant-conditioning to localise auditory, visual and spatiotemporally coincident auditory-visual stimuli throughout the frontal hemifield and to categorise temporally coincident auditory-visual stimuli according to their spatial congruency. Our results show that the integration of auditory and visual cues results in significantly more accurate and faster localisation responses and that spatial information can be used to merge different sensory stimuli and resolve conflicts between them. This confirmed the ferret as a good model to explore multisensory spatial processing.

In total, 509 single units were identified in the auditory cortex of five animals, of which 16% were influenced by visual stimulation. Our results show that modulation of activity in response to changes in auditory stimulus characteristics directly relates to task performance. For example, animals were more likely to correctly localise stimuli when activity increased in the contralateral cortex and decreased in the ipsilateral cortex. The effects of visual co-stimulation were subtle, mainly suppressive and not clearly correlated with performance on either task. A population decoding analysis showed that auditory cortical activity is, however, informative about the task context, since we were able to decode task type from trials that were otherwise equivalent in terms of the stimuli presented.

Our results support the existence of multisensory interactions in early auditory processing areas, but suggest that these areas function primarily as auditory feature detectors that operate in a task-dependent manner.

Acknowledgements

I would first and foremost like to thank my supervisors, Dr Fernando Nodal and Professor Andrew King, for their support and guidance throughout this project. In particular, Fernando for his vision, patience and incredible expertise and Andy for facilitating this opportunity and for sharing his remarkable knowledge of the field.

For her unlimited support, guidance and assistance with the histological analysis and surgeries, I am particularly grateful to Professor Victoria Bajo.

I would like to thank all those who helped with the behavioural testing, including Fernando, Vicky, Dr Joshua Gold and Susan Spires. In addition, I would also like to thank Dr Daniel Lunn for his expert guidance with the statistical analysis.

Thanks also go to the auditory neuroscience group as a whole for their insightful advice and excellent conversations. In particular, I wish to thank my lab-mates, Dr Natsumi Homma and Joshua for their friendship throughout this project.

I am particularly grateful for the support grants from Merton College, University of Oxford.

Finally, special thanks are due to my parents, Terence and Victoria, and also Arend without whose support this thesis would not have been possible.

Contributions

Chapter 1

Portions of this chapter have been submitted for publication as *King, A.J., Hammond-Kenny, A., & Nodal F.R. Multisensory Processing in the Auditory Cortex. Springer Handbook of Auditory Research.*

Chapter 2

Uni- versus Multisensory Localisation Behaviour: Hardware installed by Fernando Nodal (FN). Software written by FN. Behavioural training performed by Amy Hammond-Kenny (AHK), FN, Victoria Bajo (VB) and Susan Spires (SS). Behavioural testing performed by AHK, FN, VB, SS and Joshua Gold (JG).

Neural Activity in the Auditory Cortex during Multisensory Localisation: Surgeries performed by AHK, FN, VB. Hardware installed by AHK and FN. Software adapted from pre-implantation localisation paradigm by AHK and FN. Neural recordings performed by AHK, FN and VB.

Categorisation Behavior Based on Multisensory Binding: Software written by FN. Behavioural training performed by AHK, FN, VB and SS. Behavioural testing performed by AHK, FN, VB, SS and JG.

Neural Activity in the Auditory Cortex during Multisensory Categorisation: Software adapted from pre-implantation categorisation paradigm by AHK and FN. Neural recordings performed by AHK, FN and VB.

Brain Anatomy: Perfusion performed by AHK, FN and VB. Histological processing, and identification of recording sites performed by AHK, FN, VB and Niki Wilkinson (NW).

Chapter 3

Data processing and analysis performed by AHK using MATLAB scripts written by AHK and FN. Portions of this chapter have been published as *Hammond-Kenny, A., Bajo, V.M., King, A.J., & Nodal, F.R. (2017) Behavioural benefits of multisensory processing in ferrets. Eur. J. Neurosci., 45, 278-289.*

Chapter 4, 5 & 6

Data processing and analysis performed by AHK

Appendix A

Reconstructions of the cortex and electrode tracts made by AHK and NW. Photomicrographs in Figure A-3 prepared by VB.

Table of Contents

1 General Introduction 1

1.1 Overview 1

1.2 Multisensory Versus Auditory Brain Areas 5

1.3 Non-Auditory Inputs at Different Levels of the Auditory Pathway 11

1.4 Origins of Non-Auditory Inputs to the Auditory Cortex 17

1.5 Functional Significance of Multisensory Interactions in the Auditory Cortex 19

1.6 Thesis Overview 29

2 Methods 33

2.1 Overview 33

2.2 Animal Welfare 35

2.3 Localisation Behaviour 36

2.3.1 Apparatus and Stimuli 36

2.3.2 Head Orienting Responses 39

2.3.3 Training and General Procedure 41

2.3.4 Data Analysis 42

2.3.5 Statistical Analysis 45

2.4 Categorisation Behaviour 47

2.4.1 Behavioural Procedure 47

2.4.2 Data and Statistical Analysis 49

2.5 Surgical Implantation of Multi-Electrode Arrays 50

2.5.1 Multi-Electrode Arrays 50

2.5.2 Animal Preparation 53
2.5.3 Surgical Implantation 54
2.6 Neurophysiological Recordings during Behavioural Testing 55
2.6.1 Recording Hardware 55
2.6.2 Recording Procedure 56
2.6.3 Electrode Advancement 58
2.6.4 Behavioural Data Analysis from the Neural Recordings 60
2.6.5 Neurophysiological Data Analysis 62
2.7 Histological Analysis 67
2.7.1 Tissue Processing 67
2.7.2 Nissl and NeuN Staining 67
2.7.3 Identification of Recording Sites 69
3 Multisensory versus Unisensory Localisation Behaviour 70
3.1 Introduction 70
3.2 Results 73
3.2.1 Localisation Accuracy Varies with the Properties of the Stimulus 73
3.2.2 Reaction and Response Times 86
3.3 Discussion 95
3.4 Appendix 103
3.4.1 Statistical Analyses 103
4 Neural Activity in the Auditory Cortex during Multisensory Localisation 114
4.1 Introduction 114
4.2 Results 117
4.2.1 Task Performance During the Neural Recordings 117

4.2.2 Types of Neural Responses Evoked During the Localisation Task 122
4.2.3 Basic Response Properties of Neurons 131
4.2.4 Cross-Modal Interactions 137
4.2.5 Correlation of Neural Activity and Behavioural Outcomes 145
4.3 Discussion 152
4.4 Appendix 161
4.4.1 Supplementary Behavioural Data for the Localisation Task 161
4.4.2 Statistical Analyses 162
5 Categorisation Behaviour based on Multisensory Binding 171
5.1 Introduction 171
5.2 Results 174
5.2.1 Task Performance 174
5.2.2 Reaction and Response Times 186
5.2.3 Comparison of Localisation and Categorisation Task Performance 192
5.3 Discussion 195
5.4 Appendix 201
5.4.1 Statistical Analyses 201
6 Neural Activity in the Auditory Cortex during Multisensory Categorisation 210
6.1 Introduction 210
6.2 Results 213
6.2.1 Categorisation Task Performance 213
6.2.2 Basic Response Properties of Neurons 216
6.2.3 Effect of Stimulus Parameters on Spiking Activity 216
6.2.4 Correlation of Neural Activity and Behavioural Outcomes 223

6.2.5 Comparison of Neural Activity between Behavioural Tasks 232
6.3 Discussion 236
6.4 Appendix 242
6.4.1 Statistical Analyses 242
7 General Discussion 253
7.1 Overview 253
7.1.1 Behavioural Advantages of Multisensory Processing 254
7.1.2 Prevalence of Multisensory Interactions in the Early Auditory Cortex 255
7.1.3 Encoding of Sound Source Location by Populations of Cortical Neurons 256
7.1.4 Behavioural Relevance of Multisensory Interactions in the Early Auditory Cortex 258
7.1.5 The Use of Spatial Cues in Merging Sensory Stimuli 259
7.1.6 How the Relative Location of Stimuli Affects Multisensory Integration 260
7.1.7 Activity in the Early Auditory Cortical Processing Areas is more closely tied to the Physical Properties of the Stimulus rather than the Animal's Perception. 262
7.1.8 Conclusion 263
7.2 Future Work 264
A Appendix 268
A.1 Organisation of the ferret auditory cortex 268
A.2 Location of Electrode Arrays 271
A.3 Laminar Location of the Recordings 278
References 281

Abbreviations

A	Anterior / Audio
A1	primary auditory cortex
AAF	anterior auditory field
ADF	anterior dorsal field
AEG	anterior ectosylvian gyrus
2-AFC	two alternative forced choice
AMLS	anterior mediolateral suprasylvian sulcus bank
AV	audio-visual
AVF	anterior ventral field
BDA	biotinylated dextran amine
BF	best frequency
BW	bandwidth
CAT	categorisation
CDF	cumulative distribution functions
CF	characteristic frequency
CI	confidence interval
CTβ	Cholera toxin subunit β
Contra	contralateral
CP	cumulative probability
DAB	3,3'-diaminobenzine
DCN	dorsal cochlear nucleus
EG	ectosylvian gyrus
ETOH	Ethanol
FF	fano factor
fMRI	functional magnetic resonance imaging
FRA	frequency response areas
H	height
IC	inferior colliculus
ICx	external or lateral cortex of the IC
Ipsi	ipsilateral
Q10	bandwidth 10 dB above threshold
Q30	bandwidth 30 dB above threshold
L	length
LFPs	local field potentials
LL	lateral lemniscus
LOC	localisation
LRSS	Lateral rostral suprasylvian sulcus bank
MEG	middle ectosylvian gyrus
MGB	medial geniculate body
MGBm	medial division of the MGB MGBm
MI	mutual information
MLE	maximum likelihood estimation
MUA	multi unit activity
NeuN	Neuronal Nuclear protein

P	probability / posterior field
PAF	posterior auditory field
PVP	posterior ventral field
PB	phosphate buffer
PBS	phosphate buffer saline
PPc	caudal posterior parietal cortex
PPF	posterior pseudosylvian field
PPr	rostral posterior parietal cortex
PSF	posterior suprasylvian field
pss	pseudosylvial sulcus
PSTH	post-stimulus time histograms
RSE	redundant signal effect
S1	primary somatosensory cortex
SIII	tertiary somatosensory cortex
SC	superior colliculus
SEM	standard error of the mean
SD	standard deviation
SOC	superior olivary complex
SPL	sound pressure level
Sp	spontaneous activity
Sp5	spinal trigeminal nucleus
sss	suprasylvian sulcus
SSY	suprasylvian visual area
STS	superior temporal sulcus
Tx	triton X100
V	ventral / visual
Ve	ventral auditory area
VP	ventral posterior field
W	width
wm	white matter

1 General Introduction

1.1 Overview

Creating a unified sensory percept requires the integration of information from different sensory modalities. This process is traditionally viewed as occurring in two distinct phases in the brain. First, unisensory signals are processed by dedicated neural pathways, which are assumed to be largely independent and hierarchical. Second, once modality-specific computations have been performed, sensory information is combined and integrated in certain higher-order association areas that implement different aspects of multisensory perception. In the cortex, classical multisensory areas have been described in the frontal, parietal, and temporal lobes, where their functions are thought to range from linking multiple sensory signals with the execution of particular motor actions to the merging of communication signals provided by the eyes and ears (reviewed by Cappe *et al.*, 2012).

That sensory pathways are organised in this fashion stems from the different forms of energy that need to be detected. This necessitates the use of specialised

transduction mechanisms for converting each form of energy into neural activity and imposes constraints on the associated neural circuits in order to overcome the idiosyncrasies of each sensory modality, such as the lack of spatial information at the cochlea or the differing temporal dynamics of visual and auditory processing. Furthermore, some perceptions, for example, the colour of a flower or the pitch of someone's voice (i.e. the perceptual attribute of a sound that corresponds to its frequency), do not have obvious equivalents in other sensory modalities. Nevertheless, even for those that do, it is often possible to identify or locate an object, such as a familiar person speaking, by using more than one sense. Although this crossmodal redundancy is extremely useful for perceptual stability should one set of cues disappear – when that person stops speaking or walks outside our field of view – sensory processing most commonly occurs in a multisensory context and the simultaneous availability of information across different modalities can have profound effects on perception and behaviour. A good example of this is provided by speech perception. In order to understand the basis for this ability, it is necessary to consider not only how the brain responds to auditory information, but also the motor aspects of speech production and consequently the associated visual articulation cues. Orofacial movements during speech production provide temporally correlated cues (Chandrasekaran *et al.*, 2009) (Fig. 1-1) which, when combined with acoustic signals, improve the detection and comprehension of speech, particularly if those signals are degraded by the presence of background sounds (Sumby & Pollack, 1954; Grant & Seitz, 2000). The tendency to merge auditory-visual speech cues is further illustrated by the McGurk illusion (McGurk & MacDonald, 1976): pairing a voice articulating one syllable with a face articulating a different syllable can result in

listeners perceiving a third sound that represents a fusion of those syllables.

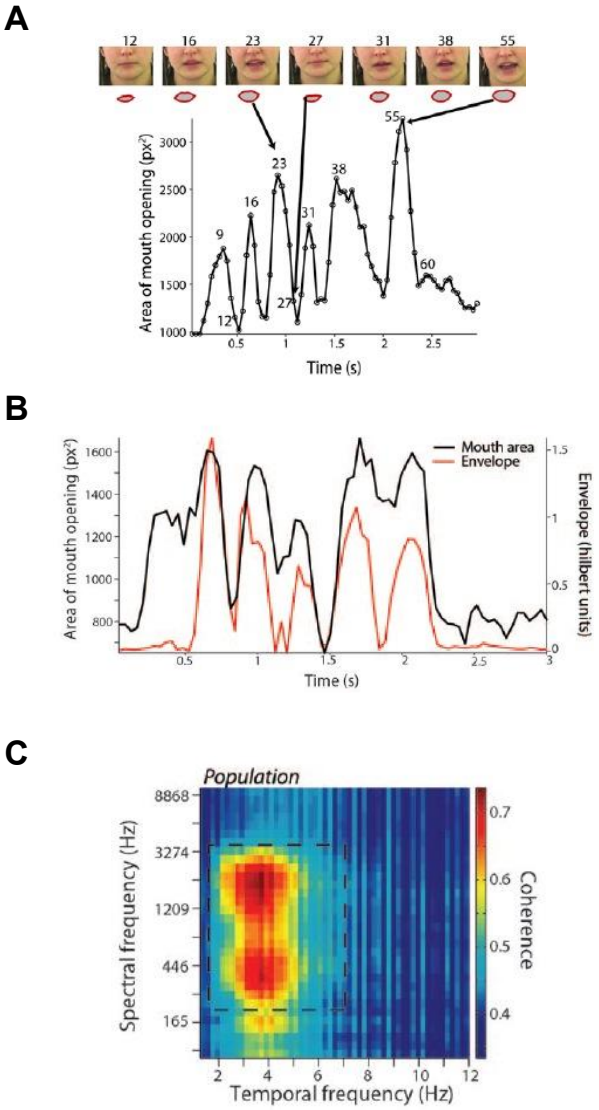


Figure 1-1. Visual and auditory statistics of human speech. (A) Example facial gestures at different frames from a video of a speaker uttering a sentence, with the red areas below each frame representing the area of the mouth opening. The graph below shows the estimated area for each mouth contour as a function of time. Arrows point to specific frames in the time series depicting different amounts of mouth opening. (B) Variation in the area of the mouth opening and the broadband auditory envelope for a single sentence from a single subject as a function of time in seconds. (C) Heat map illustrating the robust coherence between the mouth opening and auditory signal as a function of both spectral frequency band and temporal modulation frequency for 20 subjects. The rectangle drawn in dashed lines depicts the region of maximal coherence between signals. Reproduced from King, A.J., Hammond-Kenny, A., & Nodal, F.R. Multisensory Processing in the Auditory Cortex. In: Fay, R.R., & Popper, A.N, ed., *Springer Handbook of Auditory Research.* Springer (In Press). Originally adapted with permission from Chandrasekaran *et al.* (2009).

This work clearly indicates the capacity of the brain to integrate the informational content of auditory-visual speech. If the signals available in each modality are first processed independently and only subsequently combined at a specialised integration stage, one might expect the neural basis for the influence of vision on auditory speech intelligibility to reside in higher-order speech-related areas, such as the superior temporal sulcus (STS). While this is undoubtedly the case (Sekiyama *et al.*, 2003; McGettigan *et al.*, 2012), there is growing evidence that auditory and visual speech signals also interact as early as the primary auditory cortex (Schroeder *et al.*, 2008; Ghazanfar, 2009; Okada *et al.*, 2013). Furthermore, both cortical and subcortical auditory brain regions have been implicated in the various crossmodal effects that have been described for other dimensions of auditory perception. Indeed, it is a general property of sensory systems that the availability of congruent multisensory cues can result in faster responses, as well as improvements in the ability to detect, discriminate or localise stimuli (Murray & Wallace, 2012). It is therefore important to consider where and how those interactions take place, as well as the nature of the information provided by the 'non-dominant' modality, if we are to understand the impact of vision and other sensory modalities on auditory processing and perception. In this chapter, these issues will be considered in the context of the auditory pathway as a whole, but with a focus on visual influences on the auditory cortex and the implications of these effects for its primary role in hearing. Although similar questions can be asked about the functional significance of multisensory influences on processing in the visual or somatosensory cortex, the auditory cortex has been at the vanguard of research in this area. Consequently, these studies have the potential not only to improve our understanding of the computations

performed by auditory cortical neurons, but also to reveal general principles of how multisensory interactions influence perception and behaviour.

1.2 Multisensory Versus Auditory Brain Areas

Conceptually, it is difficult to classify a given brain area as unisensory if stimuli in different sensory modalities can influence the activity of the neurons found there. However, multisensory influences take different forms, ranging from a change in action potential firing in response to more than one type of sensory stimulus to crossmodal modulation of the spiking responses to one modality even if the other modality cues are by themselves ineffective in driving the neurons. In the case of the auditory cortex, there is considerable evidence for modulatory effects of non-auditory inputs on responses to sound. These interactions have been found to be particularly prevalent in functional imaging experiments, which also show that visual cues alone can activate certain parts of auditory cortex in humans (Calvert *et al.*, 1997; Pekkola *et al.*, 2005) and non-human primates (Kayser *et al.*, 2007). Similar results have been obtained using electrophysiological measurements, with local field potentials (LFPs) demonstrating widespread effects of visual or somatosensory stimuli on sound-evoked responses in both primary and secondary areas of auditory cortex (Ghazanfar *et al.*, 2005; Lakatos *et al.*, 2007; Kayser *et al.*, 2008).

Multisensory convergence in the auditory cortex appears to be more limited, however, when the spiking responses of individual neurons or small clusters of neurons are considered. This may be because LFPs (and imaging signals) also

capture the contribution of the inputs to the cortex. Again, these effects range from a change in firing rate when otherwise ineffective stimuli are paired with a sound to spiking responses evoked by visual or somatosensory stimuli (Fu *et al.*, 2003; Wallace *et al.*, 2004; Brosch *et al.*, 2005; Bizley *et al.*, 2007; Kayser *et al.*, 2008; Meredith & Allman, 2015). This apparent continuum of multisensory properties could reflect differences in the way sensory inputs converge on neurons either in the cortex (Clemo *et al.*, 2012) (Fig 1-2) or at an earlier level in the processing hierarchy.

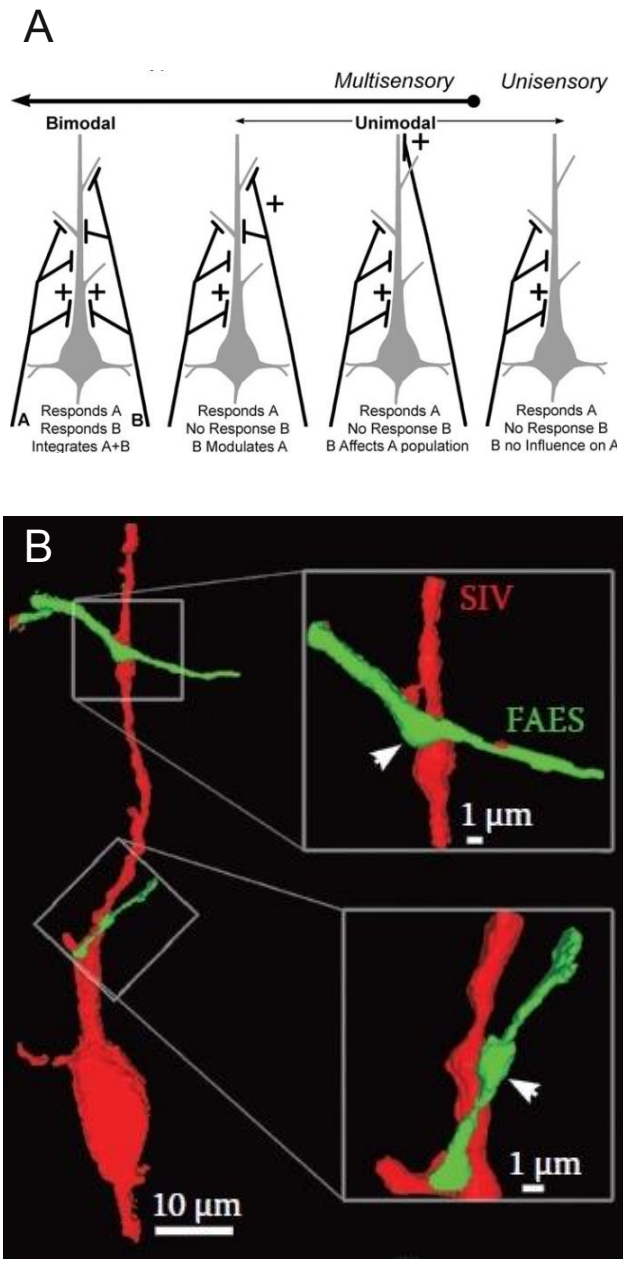


Figure 1-2. Putative patterns of synaptic connectivity underlying the range of multisensory interactions observed in the brain. (A) Neurons (grey body and dendrites) are depicted receiving afferent axons and terminals (black) from either one (at far right) or two sensory modalities ('A' and 'B'; the three cases to the left). The simplified convergence patterns vary among the multisensory neurons, so that while modality A evokes a spiking response in each case, modality B can generate a continuum of effects from producing a spiking response to modulating the response to A at the level of either the neuron or the neuronal population. (B) Confocal images of a neuron in a higher-level somatosensory cortical area (red) contacted by axons that originated in auditory FAES (green). Each axo-dendritic point of contact is enlarged on the right to show the putative bouton swelling (arrow). Reproduced from King, A.J., Hammond-Kenny, A., & Nodal, F.R. Multisensory Processing in the Auditory Cortex. In: Fay, R.R., & Popper, A.N, ed., *Springer Handbook of Auditory Research.* Springer (In Press). Originally adapted with permission from Clemo *et al.* (2012).

It is unclear what functions spiking responses to non-auditory stimuli in auditory cortex might serve, unless they convey signals that can be processed as if they were auditory in origin. In ferrets, their incidence increases between primary and high-level auditory cortical areas (Bizley & King, 2009), which likely reflects the greater density of projections to the latter from non-lemniscal thalamic nuclei (Winer & Lee, 2007) and from other sensory cortices (Bizley *et al.*, 2007). Consequently, the relative proportion of neurons that receive subthreshold modulatory inputs versus suprathreshold inputs that are capable of driving spiking activity is likely to be indicative of a progression from areas with a unisensory primary function to those more involved in merging independent inputs from different sensory modalities.

Another aspect to consider is the expected neural output of multisensory integration and to what extent it might vary in different parts of the brain. Electrophysiological recordings from multisensory neurons in the superior colliculus (SC) have led to the identification of several key principles by which different sensory inputs interact to govern the spiking activity of these neurons (King & Palmer, 1985; Meredith & Stein, 1986a, 1986b; Meredith *et al.*, 1987). The SC is characterised by the presence of topographically-aligned visual, auditory and somatosensory maps. In other words, the different modality signals arising from a particular direction – and therefore potentially from the same source – can be represented by the same neurons. The strongest enhancement of the unisensory responses of SC neurons has been shown to occur when the component stimuli are weakly effective in eliciting a response, a phenomenon that is known as the principle of inverse effectiveness, and when those stimuli occur at approximately the same

time and originate from the same region of space. By contrast, pairing strongly effective unisensory stimuli typically produces little or no enhancement, as do multisensory signals that are widely separated in time or space. Indeed, this can reduce the firing rate elicited by unisensory stimulation. That these principles operate clearly makes sense since the relative timing and location of sensory signals are important factors in determining whether they belong to the same object and should therefore be bound together, or to different objects.

Similar principles of multisensory integration have been observed in cortical neurons (Stein & Wallace, 1996) and for various behavioural tasks, including the sensory-guided orienting responses with which the SC is likely to be involved (Stein *et al.*, 1988; Bell *et al.*, 2005). However, attempts to apply them to population and sometimes indirect measures of neural activity, such as fMRI, have proved to be less straightforward (Stevenson *et al.*, 2014). Moreover, it is an oversimplification to assume that improved perceptual abilities necessarily result from increased neural activity. For one thing, neurons in the auditory cortex frequently exhibit cross-modal suppression – the response to the primary auditory stimulus is reduced when combined with visual or somatosensory stimuli (Bizley *et al.*, 2007; Kayser *et al.*, 2008) – while other studies have stressed the contribution of changes in response latency in the presence of multisensory stimuli (Chandrasekaran *et al.*, 2013).

Other studies in which single neuron recordings were made from auditory cortex have also highlighted the importance of quantifying multisensory interactions in ways that go beyond simple changes in the number of action potentials evoked. Application of information theoretic analyses to the spike discharge patterns recorded from neurons in auditory cortex has revealed that visual cues can enhance

the reliability of neural responses and hence the amount of information transmitted even if the overall firing rate either does not change or is suppressed (Bizley *et al.*, 2007; Kayser *et al.*, 2010). This finding is consistent with earlier work demonstrating that the location and identity of sounds can be encoded by the temporal firing pattern of auditory cortical neurons (Furukawa & Middlebrooks, 2002; Nelken *et al.*, 2005). Moreover, as first highlighted by the principle of inverse effectiveness in the SC, it is important to take response magnitude into account when characterising the effects of multisensory stimulation on neuronal firing patterns. Indeed, Kayser et al. (2010) showed that multisensory stimulation can have opposite effects on the magnitude and reliability of cortical responses according to how strong the responses are to sound alone, with these opposing modes of multisensory integration potentially having different functions. Thus, enhanced but less reliable responses are likely to aid the detection of near-threshold events, whereas suppressed but more reliable responses may be more relevant for discrimination in complex multisensory environments.

Population measures of auditory cortical activity in humans (Luo *et al.*, 2010; Thorne *et al.*, 2011) and animals (Lakatos *et al.*, 2007; Kayser *et al.*, 2008) indicate that non-auditory inputs can modulate the phase of low-frequency oscillatory activity in the auditory cortex. If, as a result of phase resetting, auditory inputs arrive during the excitatory phase of these oscillations, the probability of the neurons responding is thought to be increased, potentially leading to improvements in sensitivity and enhanced performance in auditory tasks. This therefore provides another way in which visual or other sensory inputs can modulate neuronal responses to accompanying sounds without necessarily evoking spiking activity in the absence of auditory stimulation. Indeed, it has even been proposed that phase resetting may

represent one of the "canonical" operating principles used by the brain to integrate different types of information (Van Atteveldt *et al.*, 2014).

Together, these findings highlight the importance of investigating multisensory interactions at multiple levels, from the activity of individual neurons to aggregate signals, including LFPs, EEG, MEG, and fMRI, and of employing the appropriate metrics in each case to quantify the magnitude and nature of integration. This is particularly important for making sense of how non-auditory inputs influence the auditory cortex without altering its fundamental role in hearing.

1.3 Non-Auditory Inputs at Different Levels of the Auditory Pathway

In considering the potential role of multisensory interactions in the auditory cortex, it is important to examine the origin of visual projections as well as their entry point into the auditory pathway. This can provide insight into the type of stimulus-related information those inputs convey and the extent to which the signals provided have already been processed and integrated by the time they reach the cortex. At the subcortical level, non-auditory inputs have been identified in most of the main relay stations of the ascending auditory pathway. The complexity of this network, which includes more levels of subcortical processing than in other sensory modalities, makes it challenging to determine the role of these inputs. Furthermore, there is currently a poor understanding of the extent to which the multisensory interactions in the auditory cortex are inherited from the thalamus, and therefore reflect bottom-up

processing, or arise from the convergence of inputs from other cortical areas.

Before discussing where visual influences are found along the auditory pathway, it is first important to consider briefly the route by which acoustic information passes from the cochlea to the cortex. Auditory nerve fibres transmit information from the cochlea to the cochlear nucleus in the brainstem, which is the first relay for the ascending auditory pathway. The cochlear nucleus comprises three subdivisions (anterolateral, ventrolateral and dorsal), within which are found several different neuron types that differ in their anatomical location, morphology, cellular physiology, synaptic inputs, and spectrotemporal response properties. The output from the cochlear nucleus takes the form of multiple, parallel ascending pathways with different targets. One of these is the superior olivary complex (SOC), where sensitivity to binaural localisation cues emerges. The various ascending tracts then innervate the nuclei of the lateral lemniscus (LL) and all converge in the inferior colliculus (IC), in the midbrain, which therefore provides a relay for the outputs from each of the brainstem auditory centres. The IC in turn delivers much of the auditory input to the SC, which as we have seen is a major site for the integration of multisensory spatial information and projects to the medial geniculate body (MGB) in the thalamus, which serves as the gateway to the auditory cortex.

Classically, the ascending auditory pathway outlined above is thought to comprise a core or lemniscal projection, which is characterised by a precise tonotopic organisation at each level from the cochlear nucleus to the primary auditory cortical fields. In addition, the extra-lemniscal projection includes parts of the IC, MGB, and a belt of auditory cortex surrounding the core tonotopic fields. Neurons in these areas tend to show broader frequency tuning than those in the lemniscal projection and

their tonotopic organisation is less well defined or even non-existent. Furthermore, they often receive inputs from other sensory modalities.

Although the functional significance of multisensory convergence within the subcortical auditory pathway is often unclear, there are instances where information from other sensory modalities makes an important contribution to the 'unisensory' role of the neurons in question. Perhaps the best example is in the dorsal cochlear nucleus (DCN). The complex excitatory and inhibitory interactions displayed by type IV neurons in the DCN of the cat allow these neurons to signal the presence of spectral notches that are generated by the directional filtering properties of the external ear (Yu & Young, 2000). Together with the finding that lesions of the pathway by which DCN projection neurons reach the IC result in impaired head orienting responses to broadband sounds (May, 2000), this points to a likely role for this nucleus in sound localisation. But cats are able to move their ears, shifting the locations at which spectral notches occur relative to the head. Consequently, information about the ongoing position of the pinnae is required in order to maintain accurate sound localisation. This appears to be provided by muscle proprioceptors located in and around the pinna of the external ear, with the DCN combining monaural acoustical cues to sound source direction with somatosensory inputs about the orientation of the pinna (Kanold & Young, 2001). More recent work in rats indicates that multisensory computations in the DCN may also help to distinguish moving sound sources from the apparent movement produced by motion of the head, suggesting that the integration of auditory and vestibular inputs helps to create a surprisingly sophisticated representation of spatial information at this early stage of auditory processing (Wigderson *et al.*, 2016).

There is also evidence to suggest that somatosensory projections to the cochlear nuclei are involved not only in sound localisation, but also in suppressing the effects of self-generated noises on the central auditory system, such as those produced by vocalising and masticating (Shore & Zhou, 2006). In support of this adaptive filter function, paired stimulation of the auditory and trigeminal nerves shows that neurons in the DCN are capable of multisensory integration and, more importantly, that the majority of multisensory interactions elicited are suppressive (Shore, 2005; Koehler & Shore, 2013). Interestingly, a related effect has also been described in tinnitus sufferers, whereby some individuals are able to modulate the loudness of their tinnitus by activating the trigeminal system using orofacial stimulation (Pinchoff *et al.*, 1998). It has been suggested that the tinnitus percept arises, at least in part, from increased spontaneous activity in the DCN (Kaltenbach, 2007). Therefore, it is conceivable that in addition to suppressing neural responses to self-generated sounds, somatosensory inputs to the DCN may reduce abnormal activity associated with phantom sounds, highlighting the therapeutic potential of harnessing somatosensory inputs to the DCN to alleviate tinnitus.

Beyond the cochlear nuclei, responses to somatosensory stimulation have also been described at other stages in the auditory pathway, including the external or lateral cortex of the IC (ICx) and the medial division of the MGB (MGBm). For both stages, this activity is mediated, at least in part, by projections from the DCN and also for the ICx by projections from the spinal trigeminal (Sp5) nucleus (Shore & Zhou, 2006). But whereas somatosensory inputs to the DCN originate principally from the pinnae, in accordance with their presumed role in the processing of spectral localisation cues, those to the IC suggest a diffuse input from the entire body. Thus,

somatosensory responses in the IC are not only inherited from the DCN and may serve to suppress the effects of self-generated noises regardless of their spatial origin.

The first responses to visual stimulation in the auditory pathway are thought to be found in the IC, and recordings in behaving monkeys suggest that the prevalence of visual influences on the neurons found there may be surprisingly high (Porter *et al.*, 2007). The IC receives a sparse input from the retina (Morin & Studholme, 2014) as well as from the visual cortex (Cooper & Young, 1976; Gao *et al.*, 2015). However, the primary source of visual input to the auditory midbrain, and potentially therefore to other parts of the auditory pathway, appears be the SC. Indeed, in ferrets, the nucleus of the brachium of the IC (Doubell *et al.*, 2000) and the ICx (Stitt *et al.*, 2015) both have reciprocal connections with the SC. This provides a source of retinotopically organised input into different regions of the IC, which likely plays an important role in coordinating and updating the alignment of maps of visual and auditory space in the SC (Doubell *et al.*, 2000; Stitt *et al.*, 2015). Potentially related to this is the finding that auditory responses in the monkey IC are modulated by changes in gaze direction (Groh *et al.*, 2001; Zwiers *et al.*, 2004). If accurate gaze shifts are to be made to auditory targets, it is essential that eye position signals are incorporated in the brain's representation of auditory space. This is well known to be the case in the SC (Jay & Sparks, 1984; Hartline *et al.*, 1995) and these findings indicate that this process most likely begins in the IC.

Beyond a role in spatial processing, non-auditory inputs to the IC could contribute to other aspects of multisensory behaviour. A single case study of a human patient with a unilateral IC lesion reported a weaker McGurk effect for audio-visual speech stimuli in the contralesional hemifield (Champoux *et al.*, 2006),

although it is unclear whether this reflects multisensory processing in the IC itself.

The thalamus is the final subcortical level in the auditory pathway at which multisensory processing occurs. In addition to inheriting non-auditory inputs via ascending projections from earlier stages in the pathway, the medial nucleus of the MGB is innervated by the spinal cord, while the dorsal nucleus and the suprageniculate nucleus receive inputs from the SC (Jones & Burton, 1974; Katoh & Benedek, 1995). An added complication when discussing multisensory processing in the thalamus is that we need to consider not only those sub-divisions comprising the auditory thalamus itself (e.g., the lemniscal ventral nucleus and the extralemniscal medial and dorsal nuclei of the MGB), but also those sub-divisions regarded as higher-order or multisensory, such as the pulvinar, which projects to and receives reciprocal projections from the auditory cortex as well as other cortical areas (de la Mothe *et al.*, 2006a). A detailed description of these projections is beyond the scope of this chapter (see Cappe at al. 2012 for a review). However, their existence is important to note, given that they provide a potential route for transferring information between different cortical areas, including those belonging to different sensory modalities (Rouiller & Welker, 2000; Sherman, 2016). Moreover, cortico-thalamo-cortical circuits can also involve the primary sensory thalamus. Thus, visual and whisker signals are combined in the ventral posterior medial region of the mouse (Allen *et al.*, 2016), while activation of the primary somatosensory cortex can alter the activity of neurons in the MGB (Lohse *et al.*, 2017).

1.4 Origins of Non-Auditory Inputs to the Auditory Cortex

The studies discussed so far show that multisensory information is incorporated at different stages along the auditory pathway, with non-auditory inputs primarily, but not exclusively, targeting extralemniscal regions. Therefore, at the cortical level, it seems reasonable to expect that non-auditory influences will be most apparent in the belt areas, and this has been confirmed by anatomical and physiological experiments in a range of species (e.g., Bizley *et al.*, 2007). In addition to its subcortical origin, however, multisensory convergence in the auditory cortex has been shown to result from inputs from other sensory as well as higher-level association cortical areas.

Anatomical tracing studies have identified direct cortico-cortical connections between different sensory areas in several species. As with subcortical levels of the auditory pathway, inputs from visual and somatosensory cortical areas are distributed primarily to non-core parts of auditory cortex, such as the caudomedial belt areas in marmosets (de la Mothe *et al.*, 2006b) and macaque monkeys (Falchier *et al.*, 2010) or the fields on the anterior and posterior ectosylvian gyri in ferrets (Bizley *et al.*, 2007). This is consistent with the greater incidence of multisensory neurons in those regions and with the often non-lemniscal nature of their auditory response properties. Nevertheless, the activity of neurons in core auditory cortex can be modulated and sometimes even driven by non-auditory inputs. In the ferret, for example, around 20% of neurons in the core areas, the primary auditory cortex and anterior auditory field, were shown to be sensitive to visual (Bizley *et al.*, 2007) or tactile (Meredith & Allman, 2015) stimulation under general anaesthesia. Although sparse projections from primary or secondary sensory areas were observed in these studies, the most

profuse retrograde labeling following tracer injections in the core auditory cortex was found in visual area 20 (Bizley *et al.*, 2007) and in the rostral suprasylvian sulcal somatosensory area (Meredith & Allman, 2015). This would suggest that core auditory areas in this species are innervated mainly by higher order visual and somatosensory cortical areas. Direct connections between A1 and other primary and secondary sensory cortices have also been described in rodents (Budinger *et al.*, 2006; Stehberg *et al.*, 2014). In marmosets, core auditory cortex is innervated by secondary somatosensory cortex and by STS (Cappe & Barone, 2005), while other studies in primates suggest that non-auditory influences on A1 most likely originate from the thalamus as well as from multisensory association areas like STS (Smiley & Falchier, 2009).

Conversely, retrograde tracer studies in primates have shown that the core and parabelt areas of the auditory cortex project to the primary visual cortex (area 17 or V1) (Falchier *et al.*, 2002). Likewise, the existence of projections from auditory to visual cortices has also been demonstrated in rodents (Campi *et al.*, 2010; Laramée *et al.*, 2011; Iurilli *et al.*, 2012).

Most of the anatomical studies have used retrograde tracer injections to reveal the origins of projections to the auditory cortex. Although this approach does not provide a clear picture of the extent and the laminar distribution of their terminal fields, it is possible to infer something about the nature of the inputs from the laminar origin of the projection neurons. Thus, feed-forward cortico-cortical projections typically originate in the supragranular layers and terminate in the granular layer 4, whereas feedback cortico-cortical projections are more likely to originate in the infragranular layers and terminate in the supragranular and infragranular layers.

Following retrograde tracer injections into A1, labelled cells are found predominantly in the infragranular layers (Cappe & Barone, 2005; Budinger *et al.*, 2006). This suggests that the core auditory cortex receives mainly feedback projections from other cortical areas, and is consistent with physiological evidence in monkeys that somatosensory inputs target the supragranular layers and have a modulatory influence on A1 activity (Lakatos *et al.*, 2007). However, feed-forward projections to the auditory cortex cannot be excluded, since several studies have also reported retrogradely-labeled cells in the supragranular layers of other cortical areas (Cappe & Barone, 2005; Budinger *et al.*, 2006; Bizley *et al.*, 2007).

The relative contributions of thalamo-cortical and cortico-cortical projections to multisensory processing in auditory cortex are poorly understood. However, Budinger et al. (2006) estimated that approximately 60% of non-auditory inputs to gerbil A1 originate subcortically, with the remaining 40% arising from other sensory or multisensory cortical areas. It is, therefore, clear that a hierarchy of multisensory processing exists within the auditory pathway and that the auditory cortex in particular is likely to be involved in various functions that depend on the integration of information across different sensory modalities.

1.5 Functional Significance of Multisensory Interactions in the Auditory Cortex

Section 1.4 discussed the possible origins of the multisensory responses that have been recorded in the auditory cortex. Because there are so many subcortical and

cortical sources of non-auditory inputs in the auditory pathway, it is challenging to pinpoint specific functions for the crossmodal influences that can be observed at the level of the cortex. Indeed, establishing a causal relationship between multisensory interactions at the neural and behavioural levels is particularly difficult, since this field of research has yet to benefit to any great degree from the experimental approaches that are now available for interrogating the functions of specific neural circuits (Olcese *et al.*, 2013; Wasserman *et al.*, 2015).

Nevertheless, insights into what those functions might be can be obtained by knowing the sources of input to particular auditory cortical areas and, of course, by measuring how the responses of the neurons change in the presence of stimuli belonging to other sensory modalities. In addition to amplifying the responses of auditory cortical neurons, particularly to relatively weak sounds, visual stimuli have been shown to induce more specific effects on the sensitivity and even the selectivity of these neurons. As discussed in Section 1.1, speech perception can be profoundly influenced by the talker's facial gestures, with studies in macaque monkeys showing that neural responses to conspecific vocalisations are enhanced when accompanied by a video-clip of an animal vocalising but not when paired with a disk presented to mimic the opening of the mouth (Ghazanfar *et al.*, 2005; Ghazanfar, 2009). Similarly, by pairing complex naturalistic audio-visual stimuli, including videos and the accompanying sounds of conspecific animals, Kayser et al. (2010) found that the information gain in the auditory cortical responses was reduced when the auditory and visual cues were no longer matched in their dynamics or semantic congruency.

These visual influences have been measured in different auditory cortical areas, including A1. The complexity of the visual information involved in

interpreting articulation cues makes it unlikely that the auditory cortex receives this information directly from early visual cortices. Instead, simultaneous recordings in the auditory cortex and STS showed that spiking activity in the auditory cortex was coordinated with the oscillatory dynamics of the STS (Ghazanfar *et al.*, 2008). Thus, in the case of communication signals, the integration of multisensory information in the auditory cortex likely depends, at least in part, on top-down inputs from this area of association cortex and probably also from other cortical areas that have been shown to be entrained by lip movements during speech (Park *et al.*, 2016).

Another area where the functional significance of crossmodal interactions in the auditory cortex is starting to become clear is sound localisation. An intact auditory cortex is required for normal spatial hearing and inactivation studies in cats suggest that this reflects the contribution of A1 plus certain higher-level auditory cortical fields, such as the posterior auditory field (PAF) (Malhotra & Lomber, 2007). Although the extent to which a division of labour exists across auditory cortex in the processing of different sound features remains controversial (Schnupp *et al.*, 2011), these findings raise the possibility that the way non-auditory stimuli influence the processing of spatial and non-spatial sound properties may be area specific.

Studies in ferrets have provided some support for this. As expected from the extensive subcortical processing that takes place in the auditory pathway, neurons across different auditory cortical fields encode both spatial and non-spatial sound features. However, neurons located in the auditory fields located on the posterior ectosylvian gyrus in this species are more sensitive to stimulus periodicity and timbre than to spatial location (Bizley *et al.*, 2009). In keeping with a potentially greater role in stimulus identification, this region receives inputs from areas 20a and 20b, which

have been implicated in visual form processing (Manger *et al.*, 2004). Conversely, neurons that are most informative about the azimuthal location of auditory, visual or paired auditory-visual stimuli are found on the anterior ectosylvian gyrus (Bizley & King, 2008), which is innervated by a extrastriate visual cortex region that has been implicated in spatial processing (Philipp *et al.*, 2006; Bizley *et al.*, 2007) (Fig. 1-3).

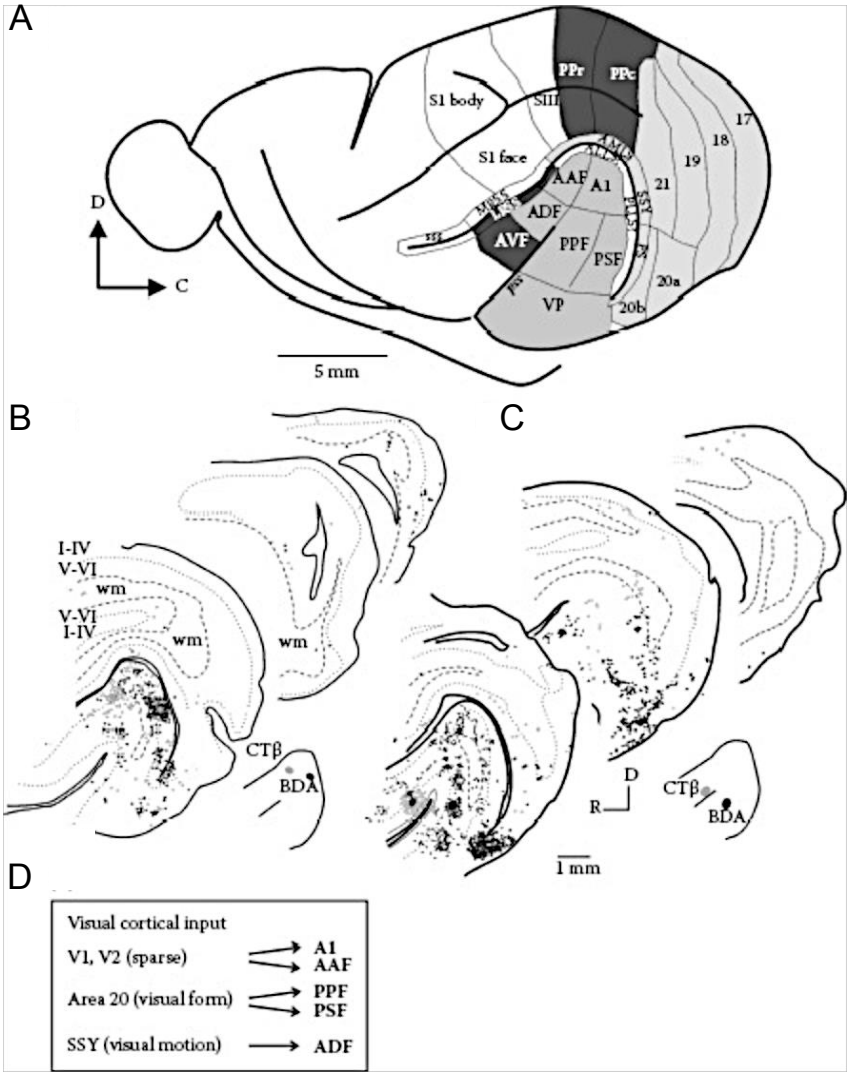


Figure 1-3. Visual inputs to ferret auditory cortex. (A) Ferret sensory cortex. Visual (areas 17–20, PS, SSY, AMLS), posterior parietal (PPr, PPc), somatosensory (S1, SIII, MRSS), and auditory areas (A1, AAF, PPF, PSF, and ADF) are shown. In addition, LRSS and AVF are multisensory regions, although many areas usually classified as modality specific also contain some multisensory neurons. (B) Location of neurons in visual cortex that project to auditory cortex. Tracer injections made into core auditory cortex (A1: biotinylated dextran amine, shown in black, and AAF: cholera toxin subunit β, shown in grey) result in retrograde labeling in early visual areas. Dotted lines delimit the cortical layers IV and V; dashed lines delimit the white matter. (C) Tracer injections made into belt auditory cortex. Retrograde labeling after CTβ injection into the anterior fields (on the borders of ADF and AVF) is shown in grey, and retrograde labelling after BDA injection into the posterior fields PPF and PSF is shown in black. Note that the extent and distribution of labeling differs after injections into the core and belt auditory cortex. Scale bars in B and C, 1 mm. (D) Summary of sources of visual cortical input to different regions of auditory cortex, with their likely functions indicated. Reproduced from King, A.J., Hammond-Kenny, A., & Nodal, F.R. Multisensory Processing in the Auditory Cortex. In: Fay, R.R., & Popper, A.N, ed., *Springer Handbook of Auditory Research.* Springer (In Press). Originally adapted with permission from Bizley and King (2012).

The interpretation of these results needs to be treated with some caution since relatively little research has so far been carried out on higher level visual or auditory cortical fields in ferrets, so a detailed understanding of the functions of these areas is not yet available. However, the crossmodal reorganisation observed following deafness is consistent with the notion that visual inputs target auditory cortical areas with related functions. Relative to hearing animals, congenitally deaf cats exhibit superior visual localisation in the peripheral field and lower movement detection thresholds (Lomber *et al.*, 2010). Cooling PAF, one of the key auditory cortical fields involved in spatial hearing, produced a selective loss of this enhanced visual localisation, whereas deactivating the dorsal zone of the auditory cortex raised visual motion detection thresholds to values typical of hearing animals. These findings therefore suggest that crossmodal plasticity occurs in cortical regions that share functions with the non-deprived sensory modality.

In keeping with the effects of matching naturalistic auditory-visual stimuli, in non-human primates, the presence of spatially-congruent visual stimuli produced an overall gain in the spatial information available in the responses of ferret auditory cortical neurons (Bizley & King, 2008). However, these effects were found to vary from neuron to neuron and the largest proportion of neurons that showed an increase in transmitted spatial information when visual and auditory stimuli were presented together was actually found in the posterior suprasylvian field, where sensitivity to sound periodicity and timbre is most pronounced.

While these studies have shown that information coding in the auditory cortex can be enhanced by the availability of matching visual cues, relatively few have measured neuronal activity while animals carry out multisensory tasks (e.g.,

Brosch *et al.*, 2005; Lakatos *et al.*, 2008; Chandrasekaran *et al.*, 2013). Therefore, the behavioural relevance of the crossmodal effects on auditory neurons observed under anaesthesia or in awake, non-behaving animals remains speculative. Moreover, where cortical recordings have been made in behaving animals, there are indications that task engagement can be accompanied by the emergence of responses to non-auditory stimuli (Brosch *et al.*, 2005; Lakatos *et al.*, 2008a) and that the modulatory nature of these stimuli may differ. Thus, visible mouth movements improve the ability of monkeys to detect vocalisations, with this behavioural advantage accompanied by shorter response latencies of auditory cortical neurons rather than changes in the magnitude or variability of their firing rates (Chandrasekaran *et al.*, 2013). This again stresses the importance of considering rate and temporal codes when investigating the impact of multisensory integration in the auditory cortex.

Measuring cortical activity during behaviour has provided other insights into the neural basis for crossmodal influences on auditory perception. Because visual information is normally more accurate and reliable, it can provide a reference for calibrating the perception of auditory space. This is particularly the case during development, when vision plays a key role in aligning the neural maps of space in the SC, as revealed by the changes produced in auditory spatial receptive fields when visual inputs are altered (King, 2009). This crossmodal plasticity compensates for growth relative changes and individual differences in the relative geometry of sense organs. But as illustrated by the ventriloquism illusion (Zwiers *et al.*, 2003), vision can also be used in adulthood to alter the perceived location of sound sources in order to resolve short-term spatial conflicts between these modalities. The neural basis for the ventriloquism illusion is poorly understood but event-related potential

and fMRI measurements have revealed changes in activity levels in the auditory cortex on trials in which participants experienced a shift in perceived sound location in the direction of a misaligned visual stimulus (Bonath *et al.*, 2007, 2014) (Fig. 1-4).

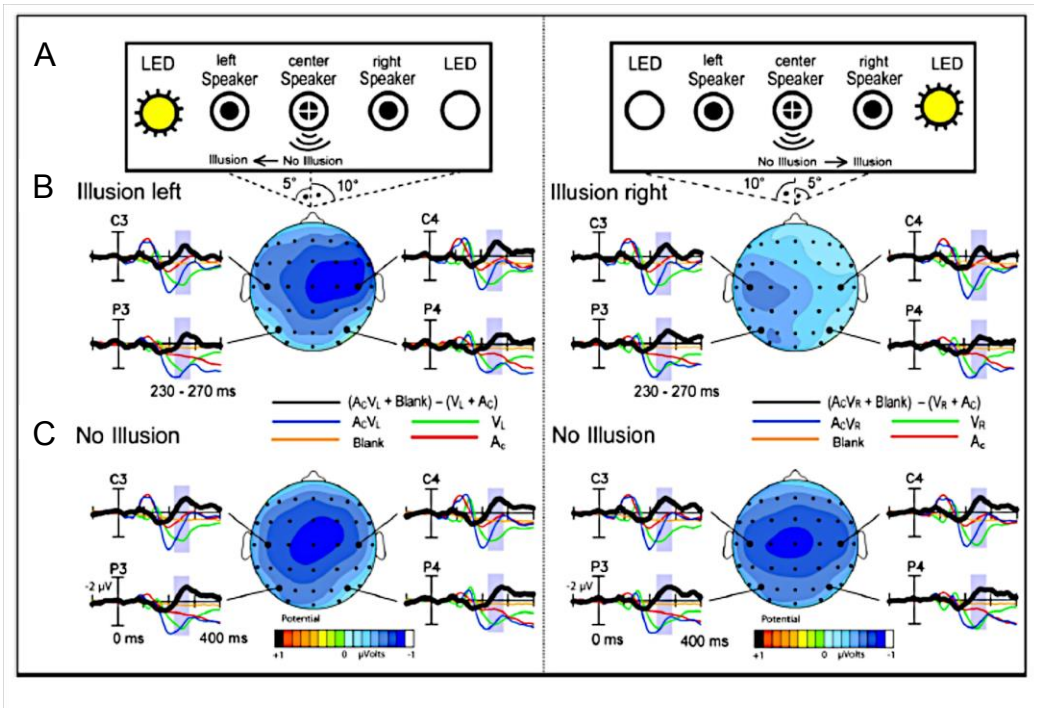


Figure 1-4. Auditory cortical correlates of the ventriloquism illusion. (A) Tones were presented from left, centre, or right speakers, either alone or in combination with flashes from light-emitting diodes on the right or left side. Left: stimulus combination of central tone (A_C) + left flash (V_L). Right: central tone (A_C) + right flash (V_R) combination. (B) Grand averaged ERPs waveforms to auditory (red), visual (green), blank (orange), and auditory-visual (blue) stimuli, together with the multisensory difference waves [(AV + blank) − (A + V)] (thick black) recorded from central (C3, C4) and parietal (P3, P4) electrodes on trials where the ventriloquist illusion was present (i.e. subjects perceived the sound as coming from the speaker on the same side as the flash). Topographical voltage maps are of the N260 component measured as mean amplitude over 230–270 ms (shaded areas) in the multisensory difference waves. Note larger amplitude contralateral to the side of flash and perceived sound. (C) Grand average ERPs and topographical voltage distributions of the N260 component on trials where the ventriloquist illusion was absent (i.e. subjects correctly reported the sound location to be at the centre). Note bilaterally symmetrical voltage distributions of N260. Reproduced from King, A.J., Hammond-Kenny, A., & Nodal, F.R. Multisensory Processing in the Auditory Cortex. In: Fay, R.R., & Popper, A.N, ed., *Springer Handbook of Auditory Research.* Springer (In Press). Originally adapted with permission from Bonath *et al*. (2012).

The growing evidence that auditory cortex may provide a substrate for visual influences on spatial hearing raises an important question. In the SC, each of the sensory representations is topographically organised and together they form overlapping maps of space. A shift in the visual map is therefore readily translated into a corresponding adjustment in the representation of auditory space by systematically retuning the neurons to a new set of spatial cue values (King, 2002). In the cortex, however, there is no map of auditory space and it is currently thought that sound source azimuth is likely to be encoded by a comparison of activity between neurons with heterogeneous spatial sensitivity within each hemisphere (Day & Delgutte, 2013; Goodman *et al.*, 2013; Keating *et al.*, 2015). Although it remains unclear how visual inputs, whether they originate subcortically or from other parts of the cortex, might "map" onto this arrangement, the finding that the ventriloquism illusion is associated with a change in the balance of activity between the left and right auditory cortical areas (Bonath *et al.*, 2014) raises testable hypotheses.

Maintaining concordant multisensory representations of space in the brain requires continuous recalibration because the spatial information provided by each modality is, at least initially, encoded using different reference frames. Thus, visual signals are encoded using eye-centred retinal coordinates, whereas auditory signals are head centred since the location of a sound source is derived from interaural time and level differences in conjunction with the spectral localisation cues generated by the head and each external ear. An important strategy used by the brain to cope with this is to incorporate information about current gaze direction into the brain's representation of auditory space. As seen in Section 1.3, this process begins in the midbrain and is widespread in monkey auditory cortex (Werner-Reiss *et al.*, 2003),

with at least some of the effects of eye position likely to arise from feedback from parietal or frontal cortex (Fu *et al.*, 2004). The importance of oculomotor information has also been demonstrated behaviourally by the surprising finding that looking towards a sound, while keeping the head still, significantly enhances the discrimination of both interaural level and time differences (Maddox *et al.*, 2014).

It is increasingly clear that focusing exclusively on the responses of neurons to the acoustic properties of sound is insufficient if we wish to understand how activity in the auditory system, and the cortex in particular, underpins perception and behaviour. As increasingly naturalistic conditions are being used to study auditory processing, more attention is being paid to the interplay between the senses. It is now known that multisensory interactions are a ubiquitous feature of the auditory system, just as they are for other sensory modalities. These interactions most commonly take the form of a modulation of auditory activity, thus other sensory inputs may provide contextual cues that signal the presence of an upcoming sound, thereby making it easier to hear. Alternatively, they may convey more specific information about the location or identity of multisensory objects and events and enhance or recalibrate the tuning properties of auditory neurons without changing their primary role in hearing.

While the application of quantitative approaches has provided valuable insights into how multisensory signals are encoded by individual auditory neurons, we currently have little understanding of the way in which populations of neurons interact to represent those signals. Moreover, given that multisensory interactions are so widespread in the brain, it remains a daunting task to decipher the specific circuits that underlie a particular behaviour. Indeed, it is becoming clear that multiple circuits exist for mediating the influence of one modality on another, as illustrated by recent

experiments in mice illustrating the different routes by which activity in the auditory cortex can suppress that in the visual cortex (Iurilli *et al.*, 2012; Song *et al.*, 2017).

The effects of multisensory processing on perception are well documented in humans, but understandably less so in other species. As we learn more about the brain regions and cell types that mediate multisensory interactions, it will be necessary to develop new behavioural paradigms to probe their role in merging different sensory stimuli and resolving conflicts between them. This will enable further assessment of the role of attention and task engagement in multisensory processing, which has so far been largely restricted to studies in primates, as well as investigation of the role of sensory experience in shaping the connections and response properties of neurons in the auditory cortex and elsewhere in the brain so that they integrate other sensory inputs that are commonly associated with sounds.

1.6 Thesis Overview

In the context of the recent work outlined in Section 1.5, suggesting that the auditory cortex may act as a substrate for visual influences on spatial hearing, in this thesis we explore multisensory spatial processing in early auditory cortical processing areas. To that intent, neural activity was recorded from the auditory cortex of ferrets, our experimental animal of choice, during the performance of two auditory-visual tasks.

The integration of information from different sensory modalities has been shown to confer wide-ranging behavioural benefits. Therefore, the first behavioural

task was designed to test the hypothesis that the presentation of combined auditory-visual stimuli results in faster and more accurate localisation responses than those elicited by the constituent unimodal sensory stimuli. A variety of factors are known to determine whether inputs from different sensory modalities are integrated including the spatial co-incidence of stimuli. Therefore, for the second behavioural task, ferrets were trained to categorise spatially congruent and incongruent auditory-visual stimuli as one object or as discrete objects, respectively, to test the hypothesis that ferrets, like other species, use spatial cues to merge sensory stimuli and resolve conflicts between them. After establishing both tasks, we next recorded neural activity from the auditory cortex during task execution. The neural recordings made during the localisation task enabled us to investigate both the extent of multisensory integration in early auditory cortical processing areas and, in particular, the hypothesis that the modulation of auditory cortical activity due to visual inputs correlates with improvements in localisation performance observed under multisensory testing conditions. In addition, the localisation task recordings provided an opportunity to investigate how sound source location is encoded by populations of cortical neurons. Likewise, neural activity was recorded whilst animals performed the categorisation task, thereby enabling us to investigate how the relative location of the stimuli affects their integration. Furthermore, the categorisation task recordings also enabled us to test whether neural activity in early auditory cortical processing areas reflects the context of the task and the animal's perception, as indicated by its behavioural choice, as well as the physical properties of the stimulus. Finally, histological analysis was performed to identify the cortical location of recording sites.

We have chosen the ferret as an animal model for a number of reasons. First, their natural inquisitiveness makes them particularly amenable for behavioural studies. This is clearly demonstrated by the range of auditory behaviours in which ferrets have been trained, including sound localisation (Kavanagh & Kelly, 1987a; Nodal *et al.*, 2008, 2010; Leach *et al.*, 2013; Keating *et al.*, 2014; Tolnai *et al.*, 2014), pitch discrimination (Walker *et al.*, 2009; Yin *et al.*, 2010) and auditory scene analysis (Ma *et al.*, 2010; Homma *et al.*, 2016). Second, they are particularly suited to auditory research because, unlike many rodent species, their audible range of frequencies (~ 20 Hz–40 kHz) overlaps that of humans. Third, their auditory cortical fields (Kelly *et al.*, 1986; Wallace & Harper, 1997; Nelken *et al.*, 2004; Bizley *et al.*, 2005) and sub-cortical as well as cortico-cortical connectivity (Bajo *et al.*, 2007; Bajo, Nodal, Bizley, *et al.*, 2010a; Bizley *et al.*, 2015) are well characterised. Fourth, the auditory cortex is on the surface of the brain, thus facilitating the targeting of the brain regions of interest with multi-electrode arrays. Finally, freely-moving ferrets have been shown to readily tolerate cranial implants (e.g. Keating *et al.*, 2014), thereby avoiding the added complication of recording from a head-fixed preparation.

Chapter 2 describes the experimental methodology employed in this thesis, including experimental design, behavioural testing, surgical implantation of multi-electrode arrays, neural recordings, and data and histological analyses.

Chapter 3 describes a novel behavioural paradigm (Hammond-Kenny *et al.*, 2017b), in the form of an operant positively reinforced multisensory localisation task, which was developed to characterise the ability of ferrets to localise auditory, visual and auditory-visual stimuli of different durations presented throughout the frontal hemifield, by measuring both their head orienting and approach-to-target responses.

Chapter 4 examines neural activity recorded from the auditory cortex, using bilaterally implanted multi-electrode arrays, during the performance of a modified version of the localisation task. Behavioural and neural data were analysed to investigate the extent of multisensory integration in early auditory cortical processing areas and, in particular, its correlation with task performance.

Chapter 5 describes a second novel behavioural paradigm, whereby ferrets were trained by operant conditioning to categorise temporally coincident auditory-visual stimuli as one object in the spatially congruent condition or as discrete objects in the spatially incongruent condition, as indicated their behavioural responses.

Chapter 6 examines neural activity recorded from the auditory cortex during the execution of a modified version of the categorisation task. Behavioural and neural data were analysed to investigate how the relative locations of the stimuli affect multisensory integration at level of the auditory cortex and, in particular, whether neural activity is more closely tied to the physical properties of the stimulus, the task context or the animal's perception, as indicated by its behavioural choice.

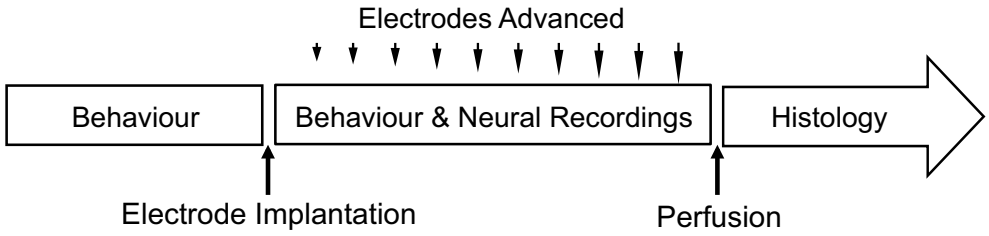
Chapter 7 provides a general discussion of the work presented in this thesis and also highlights future work suggested by our findings.

2 METHODS

2.1 Overview

Five adult female sable coated ferrets (*Mustela putorius furo,* aged approximately 6 months at training onset, sourced from Marshall BioResources, USA) were used in this study. All experiments were approved by the Oxford University Committee for Animal Care and Ethical Review and performed under licence from the United Kingdom Home Office in accordance with the Animals Scientific Procedures Act (1986, amended 2012). Ferrets were trained by positive conditioning in separate localisation and categorisation tasks using a combination of auditory and visual stimuli. Once behavioural abilities were established, multi-electrode arrays were implanted over the auditory cortex and neural activity was recorded during task performance. The experimental chronology and trajectories for individual animals are shown in Figure 2-1.

A



B

Ferret	Pre-Implantation Behaviour		Post-Implantation Behaviour/Recording		Array Location	
	Loc	Cat	Loc	Cat	Left	Right
F1305	4167 (68)	1489 (20)	3236 (52)	1496 (24)	MEG/PEG	MEG/PEG
F1324	3919 (86)	1434 (20)	7312 (143)	-	MEG	MEG/PEG
F1404	3865 (89)	1415 (23)	5968 (135)	-	MEG	MEG
F1440	3708 (83)	1374 (25)	3955 (55)	2163 (52)	MEG/PEG	MEG/PEG
F1441	4196 (53)	1426 (24)	5875 (53)	4129 (55)	MEG/PEG	PEG
Total	19855 (379)	7138 (112)	26346 (438)	7788 (131)	-	-

Figure 2-1. Experimental design. A) Experimental chronology. In each case, baseline behavioural testing was completed prior to implantation of multi-electrode arrays in the auditory cortex. Post implantation, neural activity was recorded during task performance. Following sufficient data collection, electrodes were advanced through the cortex to enable the sampling of different units during subsequent recording cycles. This procedure was repeated until no further responsive units were identified. Finally, animals were perfused trans-cardially and histological examination of the brain was performed to determine the location of recording sites. B) Individual animal experimental trajectory. Number of trials (number of testing sessions shown in brackets) completed by each animal for the localisation (Loc) and categorisation (Cat) tasks pre- and post-implantation. Array location: MEG = middle ectosylvian gyrus, AEG = anterior ectosylvian gyrus, PEG = posterior ectosylvian gyrus.

2.2 Animal Welfare

Animals were housed in groups of up to three in standard laboratory cages (L×W×H: 76.2×76.2×86.4 cm). Ferrets are known to shed their coat and, in addition, loose body weight (up to 30%) in the spring, which is co-incident with their breeding season. Therefore to mimic natural conditions, ambient conditions were varied according to British Summer Time (summer: 15:9 h light/dark cycle and 21-24°C; winter: 8:16 h light/dark cycle and 17-20°C). The cage environment was enriched with objects such as balls, tubes and shelters. In addition, at least once a week, animals were given play time outside their cages on the floor of the holding room to interact with other ferrets beyond those that they were housed with and encourage a richer repertoire of social behaviours such as "open mouth play face" (Poole, 1978).

Prior to starting the behavioural task, otoscopic examination and tympanometry were performed on each animal to exclude disease of the outer and middle ear. During behavioural testing periods, which each lasted for five consecutive days, animals were motivated to perform the task by regulating their access to water. In these testing periods, ad libitum access to dry food was provided, whereas access to water was provided only during the twice-daily (morning and afternoon) testing sessions. Animals were tested at approximately the same time each day to ensure that anticipatory effects did not vary within testing periods. If the total daily volume consumed during these testing sessions was <60 ml/kg, the typical volume consumed when ferrets have free access to water, supplementary water was provided in the form of a mash comprising ground food pellets and water to make up the deficit. Animals' body weights were measured on a daily basis and compared to

individual baseline weights recorded at the start of each testing period. In the case of an animal losing > 15% of their baseline weight, water regulation was stopped until its body weight recovered. Each testing period was followed by a break of two days over the weekend, during which the animals were provided with free water access.

All surgical procedures were performed under aseptic conditions and general anaesthesia. In addition, perisurgical analgesia was provided to minimise pain and distress (see Section 2.5 for details of the drugs and dosages used).

2.3 Localisation Behaviour

2.3.1 Apparatus and Stimuli

The localisation task was performed in a custom-built circular arena (70 cm radius) housed in a dimly illuminated (11.8 lx) sound-attenuated chamber (Fig. 2-2A). Animals were monitored from outside the chamber via closed-circuit TV monitors. To initiate a trial, animals were required to stand on a central platform (L×W×H: 30×10×2.5 cm) and nose-poke at the central waterspout, thereby ensuring that they were actively engaged with the task and facing straight ahead (defined as 0° location) when the stimulus was presented (Fig. 2-2B). Stimuli were presented from one of seven locations separated by 30° intervals in the horizontal plane around the perimeter of the frontal hemifield (±90°). At each location there was a loudspeaker (FRS 8, Visaton, Crewe, UK; LED: LTW-2S3D8, Lite-On, Milpitas, CA) and a 2.5

cm Ø translucent plastic dome illuminated by a white light emitting diode (LED, LTW-2S3D8, Lite-On) to present auditory and visual stimuli, respectively. Responses were recorded at each position by the interruption of an infrared beam located in a drinking port below each loudspeaker - LED pair. A fixed water reward (typically 150-200 µL) was provided from a spout in the drinking port if the animal correctly localised the stimulus by approaching it.

The first spout licked (approach-to-target response) and the time between the stimulus onset and this response were recorded along with the stimulus properties. Stimulus presentation, response registration and reward delivery were controlled by a personal computer communicating with a System 3 RX8 multi I/O processor (Tucker-Davis Technologies, Alachua, FL) working at a sampling rate of 100 kHz, using custom written MATLAB (MathWorks, Natick, MA) scripts.

Stimuli included auditory, visual or both combined. Auditory stimuli consisted of single presentations of broadband noise bursts (with a low-pass cut-off frequency of 30 kHz) that were generated *de novo* on each trial. To disrupt 'absolute level cues' arising from acoustic shadowing by the animal's body and thereby prevent localisation based on the relative loudness of stimuli, sound levels were roved pseudo-randomly across trials from 56-84 dB SPL in 7 dB steps. In addition, to prevent localisation based on spectral differences caused by the use of different loudspeakers, auditory stimuli were spectrally matched by convolving the signal with the respective loudspeaker's transfer filter. Visual stimuli consisted of a single continuous illumination of the translucent plastic dome, positioned immediately below each loudspeaker and at 10 cm above floor level. During multisensory trials, spatially congruent auditory and visual stimuli were presented simultaneously.

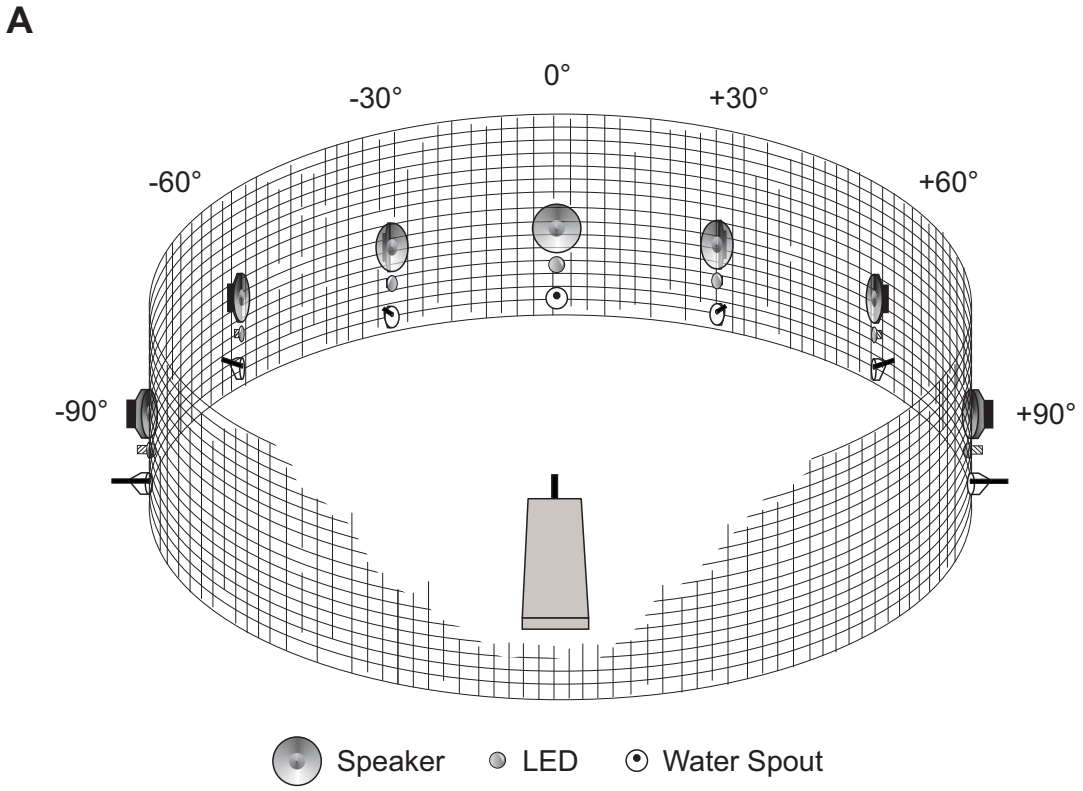


Figure 2-2. Behavioural task schematic. (A) Behavioural testing apparatus. Auditory, visual or auditory-visual stimuli were presented from one of seven loudspeaker-LED pairs located at 30º intervals around the frontal perimeter of a circular arena (radius 70 cm). Correct approach-to-target responses were rewarded with water provided from a spout located underneath each loudspeaker-LED pair. Negative and positive angles denote the locations of loudspeaker-LED-spout combinations to the left and right of the midline, respectively. (B) Schematic representation of the timing and sequence of a single trial. To initiate a trial, animals were required to stand on the central platform and lick the central spout, which triggered stimulus presentation after a variable delay (300 – 500 ms). The animal's head position was tracked for 1 s following stimulus presentation and the peripheral spout first approached by the animal was recorded.

Reaction and response times are defined as the time from stimulus onset to the initiation of the head movement and the first peripheral spout approached respectively.

2.3.2 Head Orienting Responses

In addition to the approach-to-target responses, the change in the animal's head position was recorded for the first second following stimulus onset (Fig. 2-3 A-D). Head orienting responses were measured by tracking the x-y coordinates of an adhesive reflective strip attached to the midline of the scalp at a rate of 60 frames per second, using an overhead infrared-sensitive camera (DMK 21BF04, The Imaging Source GmbH, Bremen, Germany). From these x-y co-ordinates, the head orienting reaction time and final bearing relative to the initial head position were derived using custom written scripts implemented in MATLAB. The start of the head orienting response was defined as a movement in the same direction over three consecutive frames, with the timing of the first of these frames relative to stimulus onset taken as the reaction time. The head movement was considered complete once a change in direction was detected. The final head bearing was taken as the mean angle from the last three frames of this movement or, if no change in direction was detected, as the mean angle from the last three frames recorded. Head orienting data were excluded from the subsequent analysis if the head bearing at stimulus presentation deviated by >30° from the midline or if the reaction time was >500 ms.

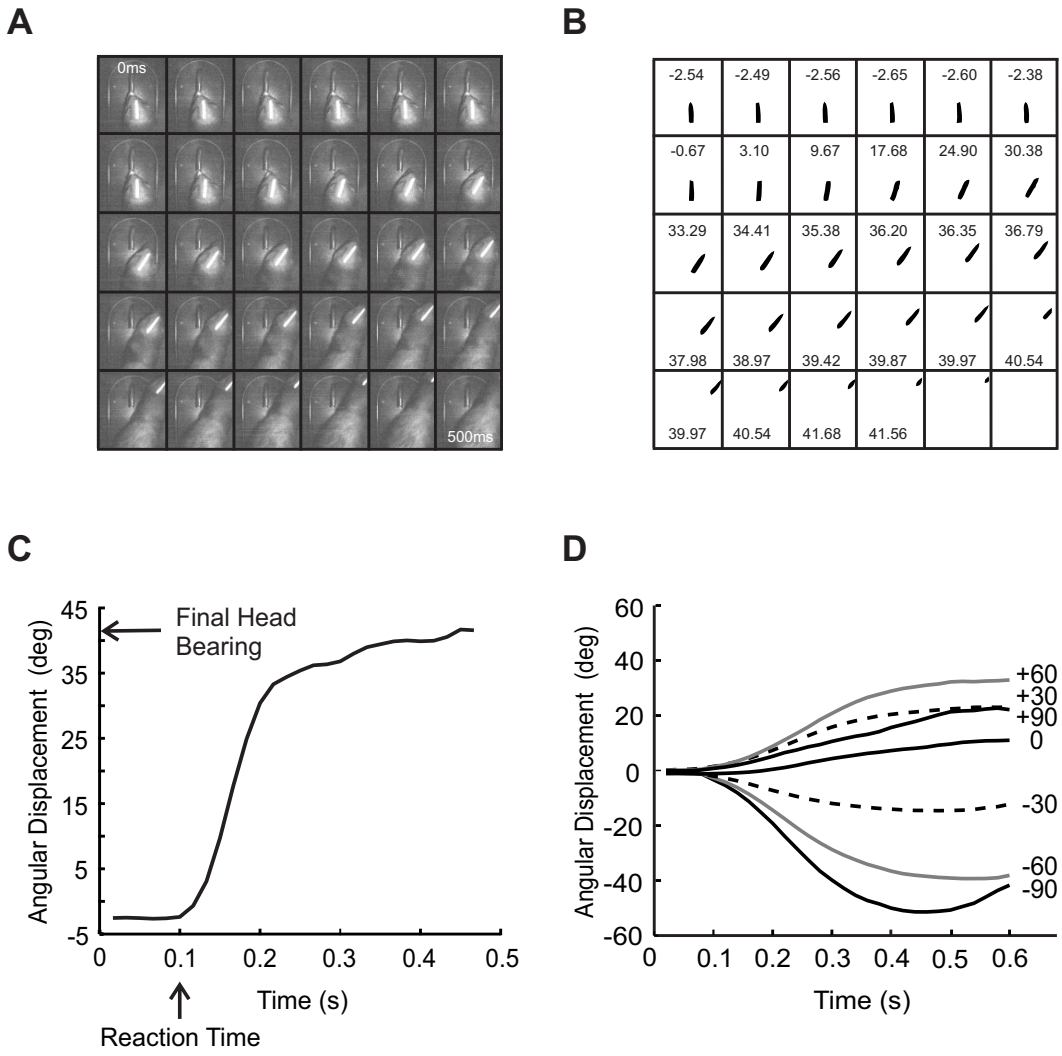


Figure 2-3. Head tracking method. (A) Stills taken by the overhead camera showing the rotational movement of the reflective strip attached to the animal's head over 500 ms after stimulus presentation. (B) Results of the automated image processing to detect headstrip orientation. (C) Plot showing how final head bearing and reaction time are derived from the angular displacement of the head for the same trial as depicted in A and B. (D) Plot showing how the angular displacement of the head varies over time for the different stimulus locations. Note only data from auditory-visual trials, in which correct approach-to-target responses were made, are shown. Negative and positive values indicate stimulus locations to the left and right of the midline, respectively.

2.3.3 Training and General Procedure

Naïve animals took approximately one week to learn the behavioural procedure. During procedural training, animals were taught to stand on the platform and lick the central waterspout for a reward. Over the course of training, the probability of this reward was gradually reduced to 1/20 trials. Once animals reliably positioned themselves correctly on the central platform, nose-poking the central spout was followed (after a variable delay of 300–500 ms) by continuous auditory-visual stimulus presentation from one of the seven pseudo-randomly chosen locations. A peripheral water reward was provided only if the animals correctly approached the location from which the stimulus was presented. All procedural training was carried out using auditory-visual stimuli to avoid introducing a training bias towards either sensory modality. Once familiar with this procedure, the ability of animals to localise auditory-visual, auditory and visual stimuli, respectively was tested in three separate blocks, in order to maximize the number of trials for each stimulus condition and ensure a stable performance. Prior to starting each block, animals were familiarised with the stimulus condition until they showed no improvement in localisation performance for stimuli of 2000 ms duration over five consecutive sessions (mean ± SEM; auditory-visual, 97 ± 1%; auditory, 92 ± 3%; visual, 83.8 ± 3%). Once criterion performance was achieved, behavioural testing was commenced and stimulus duration was progressively decreased in six steps (1000, 500, 200, 100, 40 and 20 ms), following the completion of at least 210 trials at each of the durations. To avoid a bias towards any particular stimulus location, incorrect responses were followed by at least one correction trial (same stimulus presented from the same

location) and further incorrect responses were followed by up to two easy trials (continuous stimulus presentation from the same location until a response was made). Data from correction, easy and centre-rewarded trials were not analysed.

2.3.4 Data Analysis

All data were analysed using MATLAB. The results presented are based on the analysis of 19,855 approach-to-target responses, out of which 18,267 trials yielded a head orienting trace. For the analysis, data were pooled across sound levels (56–84 dB SPL), since previous studies of auditory localisation in the ferret have identified no difference in head orienting or approach-to-target performance across this range (Nodal *et al.*, 2008). Head-orienting performance was measured by calculating final head bearings and error magnitudes (the angular difference between the final head bearing and the stimulus location). Final head bearing errors were standardised so that an undershoot resulted in a negative value and an overshoot in a positive error, regardless of whether the stimulus was presented from either the left or the right of the midline. In addition, approach-to-target performance was measured by the calculation of percentage correct scores (number of correct trials / (number of correct trials + number of incorrect trials) × 100) and error magnitudes (the difference between the target and response locations). The mutual information (MI) between the approach-to-target response location or final head bearing and the target location for the different stimulus durations and types was calculated using the formula:

$$\mathrm{MI}(r,s) = \sum_{r,s} p(r,s) \times log_2 \left[p(r,s) \div (p(r) * p(s)) \right]$$

where r is the response location or final head bearing (binned in 7.5° bins), s is the target location, MI(r,s) is the MI between r and s, p(r, s) is the joint probability of r and s and is equivalent to p(r|s)*p(s), where p(s) and p(r) are obtained from the overall distribution of target locations and response (either approach-to-target or head bearing) locations, respectively. To quantify the behavioural performance gain attributable to having information available from a second sensory modality, the percentage gain was calculated (multisensory performance – unisensory performance) / (unisensory performance) × 100). Mean and standard error of the mean were used to summarise data unless otherwise specified.

In addition, head orienting reaction times and approach-to-target response times were analysed. In order to prevent any bias due to a possible lack of attention or task engagement, which might have resulted in abnormally slow responses, reaction times > 250 ms and response times > 5 s were omitted from the analysis.

An important behavioural manifestation of multisensory integration is the speeding up of responses, a phenomenon known as the redundant signal effect (RSE; Todd, 1912). This effect can be accounted for by two classes of model (Fig. 2-4): probability summation (race model; Raab, 1962) and signal integration (co-activation model; Miller, 1982). The race model assumes that signals from different sensory modalities are processed within entirely independent parallel channels, which compete for response initiation. Consequently, the RSE occurs since the likelihood of a faster response time is greater when signals are available in more than one channel

than with one alone. In contrast, the co-activation model accounts for the RSE in terms of energy summation, whereby signal integration across channels results in the threshold for response initiation being reached more quickly.

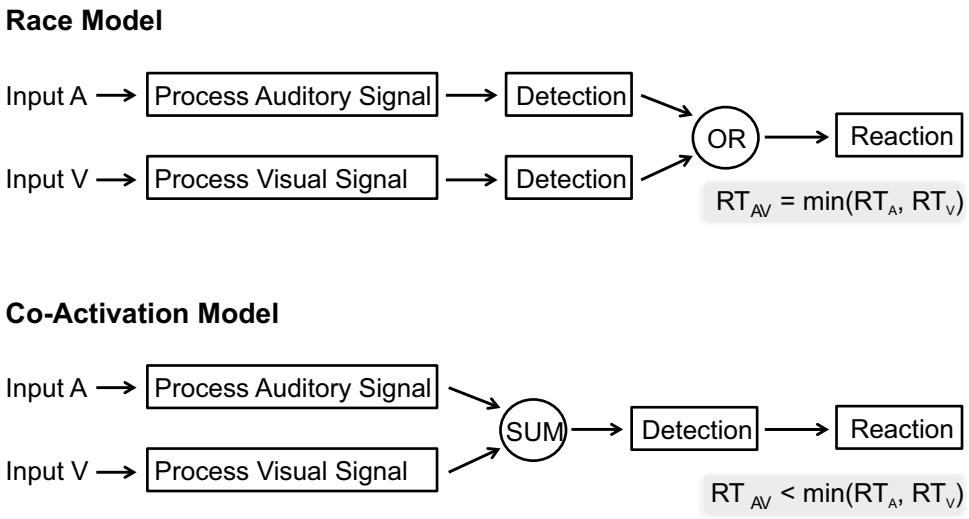


Figure 2-4. Models accounting for the redundant signal effect. (Top) The race model proposes that sensory inputs are processed within entirely independent parallel channels with the fastest channel (i.e. the winner) determining the reaction time. (Bottom) The co-activation model proposes that sensory inputs are integrated across channels, which results in the signal reaching the threshold for detection more quickly.

To identify the mechanism responsible for the RSE, reaction and response times were subjected to race model inequality analysis whereby the observed RSE is compared with the maximum facilitation expected given a race scenario, which can be derived by probability summation of the unisensory times, corrected for the assumption of independence between the unisensory channels by the inclusion of a product term in the equation (Stevenson *et al.*, 2014). If the observed RSE exceeds that predicted by probability summation then the race model is rejected and instead the existence of an integration mechanism is implied. For each animal, response

times were pooled across each stimulus condition and divided into 5% quantiles from 5% to 95% of the distribution. At each quantile, an upper limit was placed on the cumulative probability of times predicted by the race model:

$$CP_{(t)Model} = (CP_{(t)A} + CP_{(t)V}) - (CP_{(t)A} \times CP_{(t)V})$$

Where the variable $CP_{(t)Model}$ is the cumulative probability of times predicted by the race model and $CP_{(t)A}$ and $CP_{(t)V}$ are the cumulative probability of times observed for trials with auditory and visual stimuli, respectively. If at any given quantile, the observed cumulative probability of times for trials with auditory-visual stimuli ($CP_{(t)AV}$) exceeds the $CP_{(t)Model}$ value, then the race model can be rejected.

2.3.5 Statistical Analysis

Statistical analyses were performed using the R software package (www.r-project.org). The effects of stimulus modality, duration and location on head orienting and approach-to-target performance were investigated by fitting a linear mixed effects model to head orienting error data and a generalised linear mixed model for Bernoulli-distributed responses to approach-to-target response data. For both analyses, a mixed effects model was used to enable the introduction of a random effect across ferrets, which allowed for within ferret correlation. To optimise the fit of the linear mixed effects model, head orienting error values were first converted

into radians and then transformed using the sine function. The fit of the linear mixed effects model was checked using Shapiro-Wilk's test and by generating QQ-plots to test that the model residuals were normally distributed. Similarly, the fit of the generalised linear mixed effects model was checked using a Monte-Carlo simulation of the fitted model with 20 samples from a binomial distribution. Half-normal QQ-plots of the Pearson residuals for the observed and simulated data were then compared to check for a good match (for full details of the method see Collett, 2002).

The effects of stimulus modality, duration and location on reaction and response times were investigated by fitting separate linear mixed effects models. An initial inspection of the reaction and response time data showed that their distribution was positively skewed, which correlates with the results from similar operantly conditioned tasks (Leach *et al.*, 2013; Marmolejo-Ramos *et al.*, 2014; Gold *et al.*, 2015). Therefore, to optimise the fit of the models, times were first normalised using a reciprocal transformation. The fit of the models was checked using Shapiro-Wilk's tests and by generating normal QQ-plots as described above. Finally, to assess the significance of race model violations across individual animals, separate one-sided t-tests comparing the observed and modelled cumulative probability values were performed on each quantile that exhibited a group average model violation.

2.4 Categorisation Behaviour

2.4.1 Behavioural Procedure

Following completion of the localisation task, animals were trained in a second task, which required them to categorise the spatial congruency of auditory-visual stimuli (Fig. 2-5). Water access regulation, testing apparatus, stimuli generation and trial sequence were the same as described previously for the localisation task (see Section 2.3). Upon trial initiation, either a spatially congruent or incongruent auditory-visual stimulus was presented. On spatially congruent trials, unisensory stimulus components were presented from the same location (-30° or +30°) and animals were rewarded for approaching the presentation site. In contrast, on spatially incongruent trials, unisensory stimulus components were presented from separate locations (-30° and +30°) and animals were rewarded for approaching a neutral site (0°). For both spatially congruent and incongruent trial types, unisensory stimulus components were always presented simultaneously (i.e. temporally congruent). Overall, stimuli were presented pseudo-randomly with a 2:1 ratio of congruent to incongruent trials to ensure an equal probability of a reward at each response location (-30°, 0° and 30°) and, therefore, to prevent a response bias towards a particular response location.

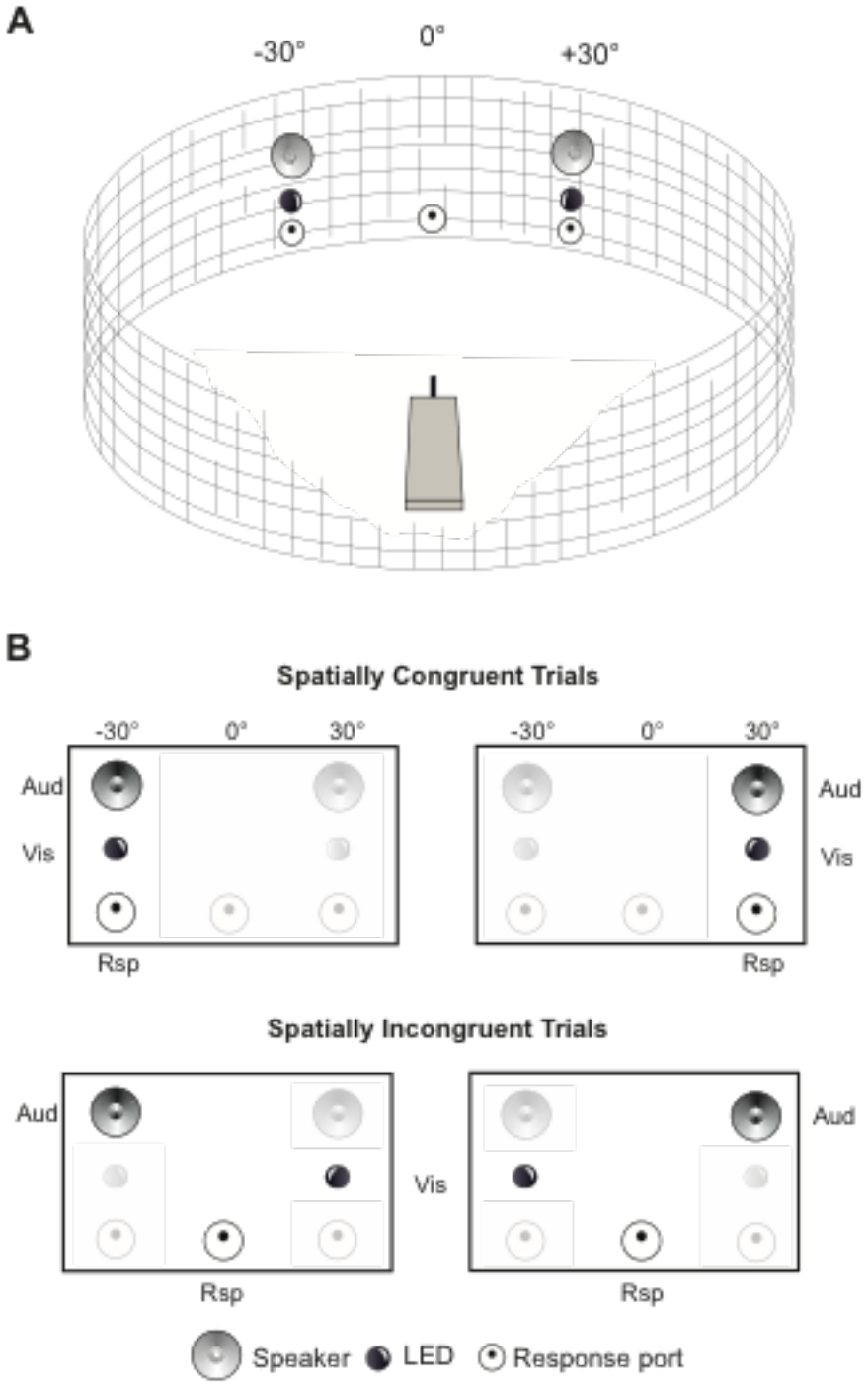


Figure 2-5. Schematic representation of the categorisation task. (Top) Diagram of the apparatus used for the categorisation task. (Bottom) Stimulus composition. On spatially congruent trials, auditory and visual stimulus components were presented from the same location (-30 or +30°) and animals were rewarded for approaching the presentation site. In contrast, on spatially incongruent trials, auditory and visual stimulus components were presented from separate locations (-30 and +30°) and animals were rewarded for approaching a neutral site (0°).

Prior to starting behavioural testing, animals were first familiarised with the categorisation task until they consistently performed >90% of trials correctly for stimuli of 2,000 ms duration. Once criterion performance was achieved, behavioural testing was commenced and stimulus duration was progressively decreased in six steps (1000, 500, 200, 100, 40 and 20 ms), following the completion of at least 210 trials at each of the durations. During each trial, the animal's head orienting reaction and approach-to-target response were recorded, as described for the localisation task. To avoid a bias towards any particular stimulus location, incorrect responses were followed by at least one correction trial (same stimulus presented from the same location) and further incorrect responses were followed by up to two easy trials (continuous stimulus presented from the same location until a response was made). Data from correction, easy and centre-rewarded trials were not analysed.

2.4.2 Data and Statistical Analysis

The results presented from the categorisation task are based on the analysis of 7138 approach-to-target responses, of which 6119 trials yielded a head orienting trace. For the analysis, data were pooled across sound levels (56 – 84 dB SPL). Head-orienting performance was measured using final head bearings, error magnitudes (the absolute angular difference between the final head bearing and the stimulus location) and reaction times. Similarly, approach-to-target performance was measured using percentage correct scores (number of correct trials / (number of correct trials + number of incorrect trials) × 100) and response times. As for the localisation task analysis, statistical tests were performed by fitting separate linear mixed effects

models to the head orienting error, head orienting reaction time and approach-to-target response time data and a generalised linear mixed effects model for Bernoulli distributed responses to the approach-to-target response correctness data. The goodness-of-fit of each model was checked as described in Section 2.3.5.

2.5 Surgical Implantation of Multi-Electrode Arrays

2.5.1 Multi-Electrode Arrays

Following the completion of baseline behavioural testing, Warp 16 multi-electrode arrays (Neuralynx Inc, Bozeman, MT, USA), comprising a 4×4 matrix of stainless steel electrode guide tubes (30 G) embedded in epoxy resin, circuit board and 16 pin omnetics connector, were implanted bilaterally over the auditory cortex (Fig. 2-6). Arrays were small (H×W×L: 17×12×8 mm, footprint on the cortex surface 4×4 mm) and light (0.8 g) and, therefore, ideally suited for the ferret. Prior to implantation, each guide tube was loaded with a 2 MΩ tungsten electrode (FHC, Bowdoinham, ME, USA). The electrode was held in the guide tube by a bend in its un-insulated tail segment. This configuration allowed for the conduction of electrical signals from the electrode to the guide tube and also for the electrode to be advanced through the cortex. Prior to loading, the electrode was trimmed to 20 mm (5 mm longer than the guide tube) and the insulation was stripped manually from its tail segment. Under microscopic visualisation, each electrode was inserted tail segment first into the guide tube, bent by ~30° to ensure electrical contact, trimmed and pushed back until

the electrode was flush within the guide tube (for full details of the loading procedure see Eliades & Wang, 2008; Roy & Wang, 2012). Following insertion of all electrodes, a ground wire was soldered to ground contact pins on the array.

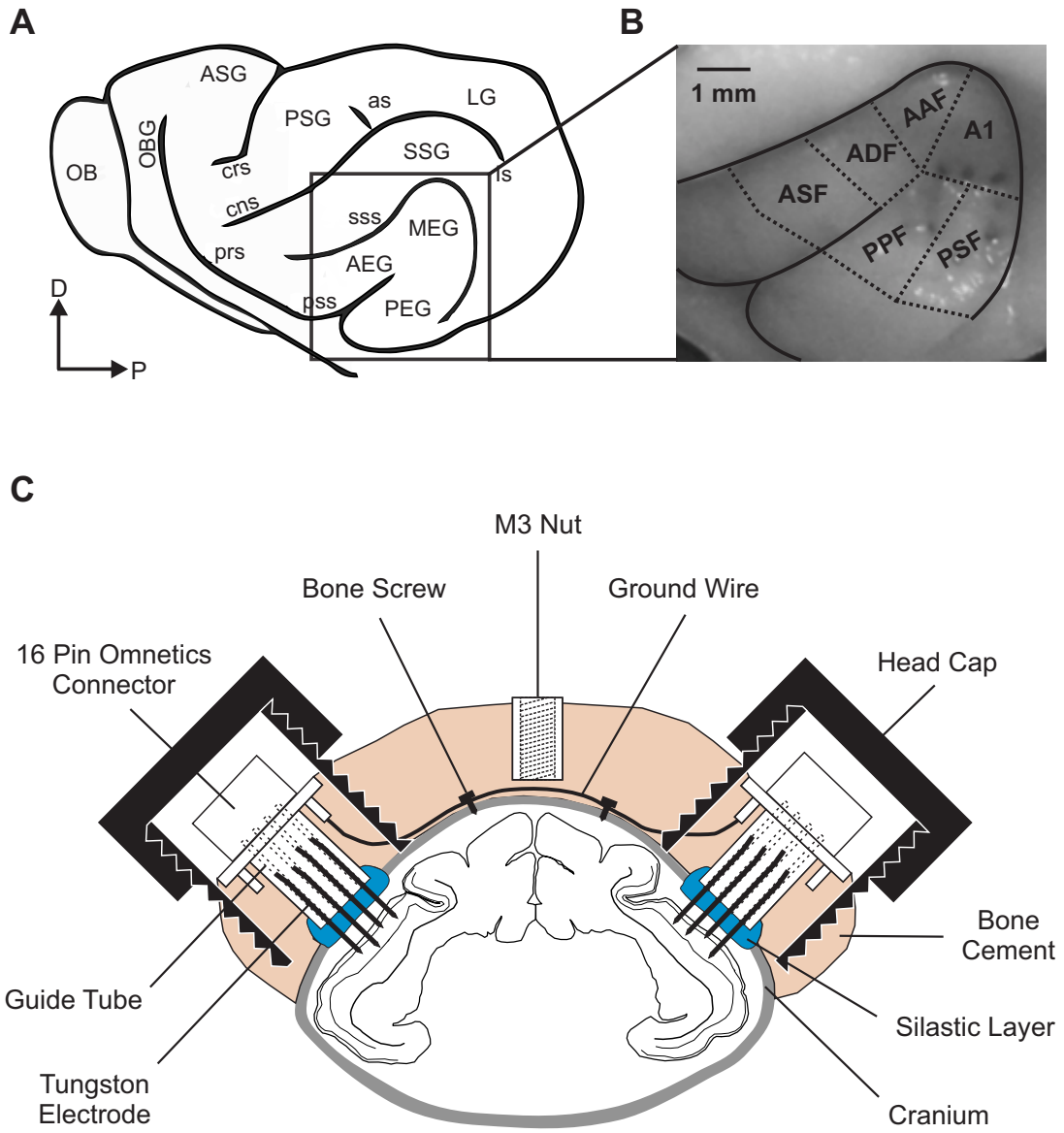


Figure 2-6. Multi-electrode array implantation in the ferret auditory cortex. A) Schematic of the ferret brain showing the principle gyri and sulci, with the auditory cortex located on the ectosylvian gyrus highlighted. B) Photograph of the right surface of the ectosylvian gyrus showing electrode imprints from an implanted array. A schematic of the auditory fields is superimposed. C) Cross-sectional diagram of the cranial implant comprising two multi-electrode arrays with protective head caps implanted bilaterally over the auditory cortices. Abbreviations: as, anseate sulcus; A1, primary auditory cortex; AAF, anterior auditory field; ADF, anterior dorsal field; AVF, anterior ventral field; ASG, anterior sygmoid gyrus; cns, coronal sulcus; crs, cruciate sulcus; D, dorsal; LG, lateral gyrus; ls, lateral sulcus; OB, olfactory bulb; OBG, orbital gyrus; P, posterior; PPF, posterior pseudosylvian field; prs, presylvian sulcus; PSF, posterior suprasylvian field; PSG, posterior sygmoid gyrus; pss, pseudosylvian sulcus; SSG, syprasylvian gyrus; sss, suprasylvian sulcus; posterior field.

2.5.2 Animal Preparation

General anaesthesia was induced with a mixture of metetomidine (Domitor, 0.022 mg/kg, IM, Vetoquinol) and ketamine (Narketan-10, 0.05 mg/kg, IM, Pfizer). After induction, lidocaine spray (5%, Biorex) was administered to the glottis and the trachea was intubated. Anaesthesia was then maintained with isoflurane (IsoFlo, 2% in Oxygen, INH, Abbott) and, once stable, atipamezole hydrochloride (Antisedan, 0.5 mg/kg, SC, Jansen Animal Health) was administered to reverse the effects of metetomidine. Peri-operatively, atropine sulphate (Atrocare, 0.06 mg/kg, SC, AnimalCare) and doxapram hydrochloride (Dopram, 4 mg/kg, SC, Jansen Animal Health) were administered to reduce pulmonary secretions and regulate cardiac rhythm, respectively. In addition, buprenorphine hydrochloride (Vetergesic, 0.03 mg/kg, SC, Alstol Animal Health) and meloxicam (Metacam, 0.2 mg/kg, S/C, Boehringer Ingelheim) were administered to provide analgesia and co-amoxiclav (Augmentin X mg/kg, IV Pfizer) to reduce the risk of infection. Throughout surgery, 5% dextrose was infused to maintain hydration and viscotears were applied frequently to protect the eyes from drying (Carbomer, topical, Novartis Pharmaceuticals). The ECG, respiration rate, end-tidal CO_2 and temperature were each monitored continuously and the animal's body temperature was maintained at 38°C with an electrically heated plate and warm air circulating Bair Hugger system.

2.5.3 Surgical Implantation

Multi-electrode arrays were implanted bilaterally in the auditory cortex of animals under aseptic conditions. The animal's head was placed in a stereotaxic frame and the scalp was incised along the midline. The temporal muscles were partially excised bilaterally and the remaining muscle was retracted with the skin to expose the cranium. Bilateral craniotomies (5 mm × 5 mm) were performed over the auditory cortex at the ectosylvian gyrus. The underlying dura was left intact and covered with a thin layer of silastic (QWIK-SIL, WPI, Sarasota FL, USA) to prevent cerebrospinal fluid leakage into the guide tubes of the electrode arrays. The location of the arrays was determined intra-operatively from the sulcal pattern on the cortical surface and the local vasculature. Each array was positioned perpendicular to the cortical surface and lowered, using a micromanipulator, into the silastic layer. Once positioned, the ground wires from the array were wrapped around the exposed heads of several bone screws placed through the cranium adjacent to the craniotomy. The bone screws, ground wires and exposed cranium were sealed with dental glue (C&B Superbond, Sun Medical) to improve adhesion between the cranium and the pedestal constructed around the arrays using gentamycin impregnated bone cement (Depuy, Johnson & Johnson). A threaded plastic cylinder and removable head cap were placed over each array to provide protection, whilst enabling access to the electrodes and omnetics connector. Two nuts (metric 3 mm M3) were also embedded in the midline of the pedestal to enable head fixation during the electrode advancement procedure.

Post-operatively, animals were allowed at least four weeks to recover before restarting behavioural testing, to ensure that the interface between the skin and the

implant was stable during the subsequent recordings and also completely healed to minimise the risk of infection. For the first five days, Meloxicam (Metacam, 0.2 mg/kg, S/C, Boehringer Ingelheim), Buprenorphine (Vetergesic, 0.03 mg/kg, SC, Alstol Animal Health) and Co-Amoxiclav (Synulox, 8.75/1.75 mg/kg, SC, Zoetis) were administered to provide analgesia and reduce the risk of infection, respectively.

2.6 Neurophysiological Recordings during Behavioural Testing

2.6.1 Recording Hardware

For the neural recordings, animals were tethered by two extensible head-stage cables. At the animal's end of the cable, a digital head-stage clip (ZC16, ZIF-clip 16 Channel Digital Head-Stage, Tucker-Davis Technologies) was attached via an adaptor (ZCA-OMN16, ZIF-clip 16 Channel Adaptor, Tucker-Davis Technologies) to the omnetics connector on the implanted array. In order to prevent strain on this connection, the cable was secured by a custom-made plastic restrainer screwed onto the threaded cylinder encasing each array. At the opposite end, the cable was attached to a motorised commutator (ACO32, Tucker-Davis Technologies) mounted above the centre platform. During each trial, the commutator was disabled from 250 ms before the stimulus onset until the approach-to-target response to prevent noise from the commutator motor interfering with the electrophysiological recordings

Voltage signals from the electrodes were digitised at the head-stage

(sampling rate 25 kHz, band-pass filtered 300 – 3000 Hz) and routed via a pre-amplifier (PZ4-2 Digital Head-stage Manifold, Tucker-Davis Technologies) to a System 3 processor (RZ5D, Tucker-Davis Technologies). The interface for data acquisition was custom designed using OpenEx and RPvdsEx software (Tucker-Davis Technologies), alongside custom written MATLAB scripts to synchronise the recordings with the stimulus presentation and behavioural responses. All channels were visualised in real-time throughout each recording session to check the quality of the recordings and, therefore, ensure the absence of avoidable noise artefacts.

2.6.2 Recording Procedure

Following recovery, animals underwent a period of behavioural testing to ensure behavioural consistency with pre-implantation performance and familiarisation with tethering prior to starting the neural recordings. Once stable behavioural performance was achieved, neural activity was recorded during the performance of modified versions of both the localisation and categorisation tasks described above (see Sections 2.3 and 2.4, respectively). For both tasks, the number of possible stimulus parameter combinations was reduced to enable data collection from sufficient repetitions of each stimulus combination within the experimental timeframe. Based on the pre-implantation performance and to avoid a ceiling effect on the neural responses due to near perfect performance, stimulus duration was limited to 100 ms and sound levels were roved pseudo-randomly across a lower range from 20-55 dB SPL in 7 dB steps, whilst visual stimulus intensity remained constant across trials at

17 cd. For the localisation task, auditory-visual, auditory and visual stimuli were presented pseudo-randomly from one of two (-30° or +30°) locations.

Neural recordings were generally conducted in fortnightly cycles (Fig. 2-7). At the start of each cycle, a spike-sorting template was generated, which was then applied to all subsequent recording sessions within a given cycle to avoid spike-sorting variability from session to session. For the first 5 days of the cycle, neural activity was recorded during the twice-daily performance of the localisation task, whereas for the last 5 days of the cycle, neural activity was recorded during the twice-daily performance of the categorisation task. The order of the behavioural tasks was reversed every cycle to minimise disruption for the animals. A break of 2 days was provided between tasks. At the end of the cycle, to characterise the frequency response properties of units, animals were sedated with Metetomidine (Domitor, 0.022 mg/kg, IM, Vetoquinol) and neural activity was recorded during passive listening to pure tones of varying frequency and intensity (100 ms; 0.5-32 kHz in 1/6 octave steps; 20-90 dB SPL in 10 dB steps) presented pseudo-randomly at a rate of 1 Hz. Following the completion of the passive recordings, electrodes were advanced 100 μm further into the cortex. This procedure was repeated over several months until no further acoustically driven-spiking activity was detectable.

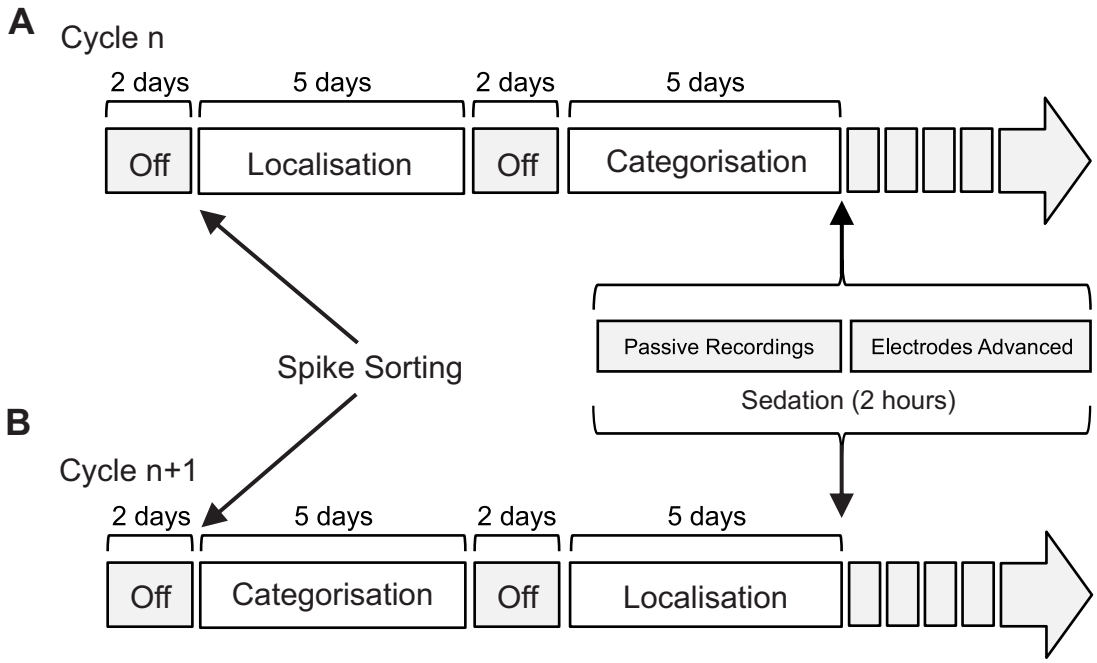


Figure 2-7. Recording schedule. Neurophysiological recordings during behavioural testing were conducted over a 14-day cycle. Neural activity was recorded during the twice-daily performance of the localisation task for 5 days and then during the twice-daily performance of the categorisation task for a further 5 days (A). Following each task, a break of two days was provided. At the end of each cycle, animals were sedated and neural activity was recorded during the passive listening to pure tones. Electrodes were then advanced into the cortex to enable the sampling of different units during the next cycle in which the order of the tasks was reversed (B). This process was repeated until no further acoustically driven spiking activity was recordable.

2.6.3 Electrode Advancement

Electrodes in the Warp-16 array were independently moveable. Therefore, following sufficient data collection, electrodes were advanced into the cortex to allow for the sampling of further units (Fig. 2-8). For the electrode advancement procedure, animals were sedated with Metetomidine (Domitor, 0.022 mg/kg, IM, Vetoquinol) and their heads were restrained in a custom frame using the two M3 nuts embedded in the implant. The animal's body temperature was monitored continuously and

maintained at 38°C with an electrically heated blanket. Electrodes were individually advanced 100 µm using an electrode pusher (Neuralynx) comprising a modified micron scale caliper, which was attached to the frame with a micromanipulator. The electrode pusher was positioned sequentially over each electrode guide tube and the probe wire was lowered into the tube to advance the electrode. Once all electrodes had been advanced, sedation was reversed with Atipamezole (Antisedan, 0.5 mg/kg, SC, Jansen Animal Health) and neural recordings were recommenced after a two-day break to allow for the settling of brain tissue around the electrodes.

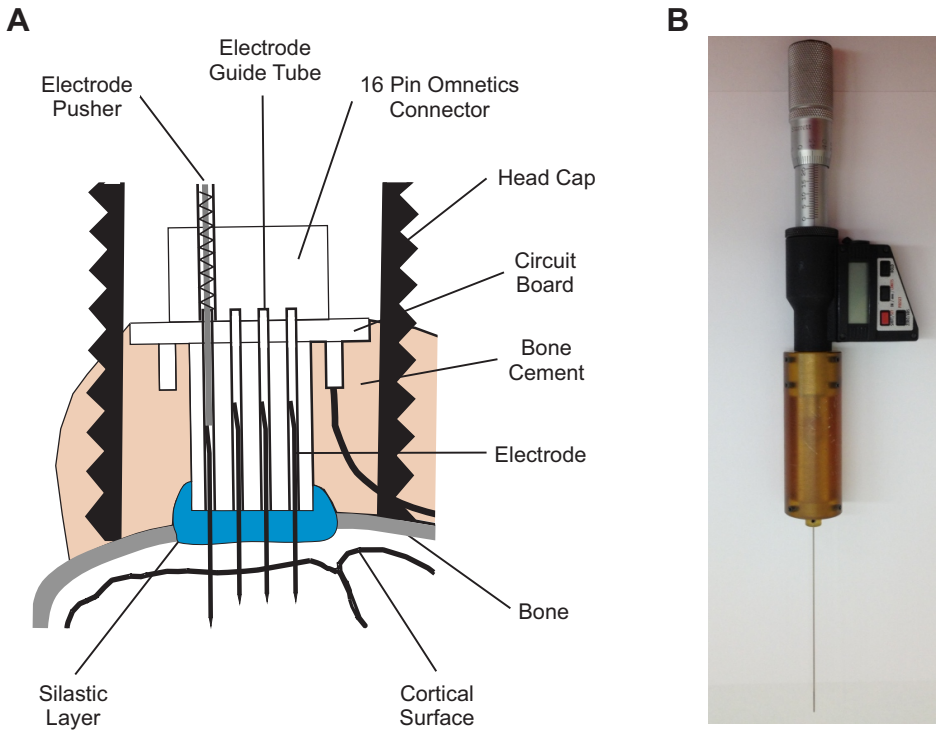


Figure 2-8. Electrode advancement. A) Illustration of the electrode advancement method. Under sedation and with head fixation, the end of the electrode pusher was positioned over the electrode guide tube until the tip was touching the array circuit board. The probe wire was then lowered to advance the electrode deeper into the brain. This procedure was repeated until all electrodes were advanced by at least 100 µm. B) Photograph of the electrode pusher, comprising a modified micron-scale caliper.

2.6.4 Behavioural Data Analysis from the Neural Recordings

For both the localisation and categorisation tasks, analysis of behavioural data from the neural recording experiments focused on approach-to-target performance. Head orienting data and, in addition, approach-to-target response times were omitted from the analysis due to the mobility changes introduced by the presence of the implants and by the fact that the animals were tethered. Statistical analysis to investigate the effects of stimulus condition, duration and location, were performed by fitting a generalised linear mixed model for Bernoulli distributed responses to the raw approach-to-target response correctness data. The model fits were confirmed by comparing half-normal QQ plots of the Pearson residuals for the observed and simulated data (see Section 2.3.5 for method).

In addition, for the localisation task, psychometric curves relating the probability of a right-sided (+30°) approach-to-target response to stimulus intensity and location were generated by fitting separate cumulative normal distribution functions to the data from auditory-visual and auditory trials. A modelled psychometric curve was also generated, to investigate whether ferrets integrate information across modalities in a statistically optimal manner, using a maximum likelihood estimator (MLE) approach (model adapted from Ernst *et al.*, 2004; Hollensteiner *et al.*, 2015). This approach sums unisensory stimulus estimates according to a weighting scheme, whereby weights are inversely proportional to the variance of the respective stimulus estimate (i.e. greater weights are given to more reliable sources). Specifically, the response probabilities for auditory and visual stimuli were used to derive a statistically optimal auditory-visual probability

according to the following method. First, for each of the sound intensities tested, a standard deviation σ_A was derived from the auditory probability (P_A):

$$\sigma_A = \frac{1}{inv.erf\ (P_A)}$$

where inv.erf denotes the inverse error function. Similarly, a standard deviation (σ_v) was derived from the visual probability (P_V) using the same equation (note that this step was performed only once since the visual intensity was kept constant). Next, the auditory and visual standard deviations were combined to obtain a modelled auditory-visual standard deviation (σ_{model}):

$$\sigma_{model} = \frac{\sigma_A * \sigma_V}{\sqrt{\sigma_A^2 + \sigma_V^2}}$$

From the modelled standard deviation a modelled probability (P_{model}) was calculated:

$$P_{model} = \text{erf}\left(\frac{1}{\sigma_{model}}\right)$$

where erf denotes the error function and σ_{model} the modelled standard deviation. Finally, a cumulative normal distribution function was fitted to the P_{model} values and the resulting curve was compared to the empirical auditory-visual curve.

2.6.5 Neurophysiological Data Analysis

Waveforms from individual units were identified and resolved online using an automated Bayesian clustering method (SpikePac, OpenEx software, Tucker-Davis Technologies, Fig. 2-9). In addition, spike sorting was supervised off line by the inspection of inter-spike interval histograms to ensure that automatically detected clusters had a clear refractory period and were therefore correctly classified as single units. The spike timing data were exported for further analysis in MATLAB, using custom made scripts. The spike-sorting template was generated at the start of each cycle and then applied to all sessions within a given cycle to ensure that clustering properties were kept constant and also to exclude activity from units that showed a neuronal drift and became unresponsive during the recording cycle.

For each unit, responsiveness to the different stimuli was established by fitting a generalised linear model for a Poisson distribution to the data obtained from the lateralisation task. For the model, three response windows were defined: onset (0-50 ms after stimulus onset), sustained (50-100 ms after stimulus onset) and offset (100-150 ms after stimulus onset). Units were classified as responsive to auditory and/or visual stimuli if the activity elicited by the respective stimulus in at least one response window differed significantly ($P<0.01$) from the spontaneous activity (measured from a 50 ms window before the stimulus onset). For each unit classified as responsive to only one modality according to the criteria above, a further comparison was performed within the driven modality to identify whether its response was significantly ($P<0.01$) modulated by the simultaneous presence of the second modality, which when presented independently failed to elicit a response. In

summary, units were classified as unimodal auditory, unimodal visual, bimodal suprathreshold or bimodal subthreshold. Unimodal neurons were defined as those that responded to only one stimulus modality (either auditory or visual). Bimodal suprathreshold neurons were defined as those that responded to both modalities when presented independently. Bimodal subthreshold units were defined as those that responded to only one modality when presented independently but exhibited significant modulation of their responses, which could take the form of either suppression or facilitation, due to the addition of the second independently ineffective modality.

Raster plots were generated for individual units, whereby each line of the plot displays the spike times, as dots, for a single trial. Pooled post-stimulus time histograms (PSTH) were also generated by summing the spike counts across trials in 2 ms bins. Minimum response latencies were calculated from the PSTH containing the responses for the relevant stimulus. Specifically, minimum response latencies were calculated as the time at which the pooled response first exceeded a criterion level, which was defined as the mean spontaneous activity plus 2.5 times the standard deviation. In addition, mean and first spike response latencies were calculated, which were defined as the mean latency of all spikes within the response window and the latency of the first spike within the response windows, respectively.

The stimulus duration (100 ms) and the precise temporal pattern of evoked activity determined the time window chosen for the analysis of spiking activity. Although the majority of units exhibited an onset response (mean auditory response latency: 27 ms), some units exhibited an offset response with a similar latency to the onset response from the stimulus offset. Therefore to capture both onset and offset

responses, a 150 ms window was used for the analysis of spiking activity.

The magnitude of cross-modal interactions exhibited by units was quantified using a modulation index (*ii*) (King & Palmer, 1985; Stevenson *et al.*, 2014).

$$ii = \frac{AV - A}{A} \times 100$$

where the variables AV and A represent the mean number of spikes elicited in a 150 ms window following auditory-visual and auditory stimulation, respectively. An increase in firing rate for the auditory-visual condition as compared to the auditory condition results in a positive *ii* and indicates response enhancement, whereas a decrease in firing rate results in a negative *ii* and indicates response suppression.

The effects of a variety of factors on the number of spikes evoked per trial were investigated by fitting separate negative binomial generalised linear mixed models to the spike count data obtained from the lateralisation and categorisation tasks. Mixed effects modeling enabled the contribution of fixed effects (i.e. stimulus condition, stimulus intensity, stimulus location, unit classification and spontaneous activity) to be established, while controlling for random effects (i.e. cell identity nested within ferret identity). A negative binomial model was required, since preliminary inspection of the data showed that the variance of the spike counts exceeded the mean and hence the data were over dispersed. In addition, the effects of spontaneous and evoked activity on the probability of a correct approach-to-target response were investigated by fitting separate generalised linear mixed models for

Bernoulli-distributed responses to approach-to-target data obtained from both tasks.

A neural decoding analysis (method adapted from Meyers, 2013) was also employed to analyse the neural activity recorded from the auditory cortex during the categorisation task. This approach uses a machine learning technique to study the information contained in the neural spike train. Specifically, it was used to investigate whether the neural activity recorded from the auditory cortex encodes sufficient information to identify the experimental conditions (e.g. stimulus location, intensity or type) or the animal's behavioural response (e.g. response correctness). First, the data were split into two parts. A pattern classifier was trained using the first smaller part of the data (the training set) to learn the relationship between the spike trains and the experimental conditions (or the behavioural response). Next, the classifier was used to predict the experimental condition (or the behavioural response) from the neural activity for the trials in the second larger part of the data (the testing set). The pattern classifier employed for the analysis used a maximum correlation coefficient method. This approach chooses the scenario as the prediction, which shows the highest Pearson's correlation coefficient between the information extracted from the spike pattern of a particular trial and bench marks calculated from the training set.

Frequency response areas (FRA) for pure tones were generated by summing activity that occurred over 100 ms time window during stimulus presentation across all presentations (10 repeats) for each frequency-level combination. The resulting spike matrix was smoothed using a nine-point running hamming window (Bizley *et al.*, 2005). From the smoothed matrix, a tuning curve was constructed by applying a responsiveness criterion, defined as the mean spontaneous rate plus 20% of the peak

firing rate (Sutter & Schreiner, 1991; Bizley *et al.*, 2005). FRAs were inspected to ensure accordance between the spike matrix and tuning curve. Several parameters were derived from the FRA including best frequency (BF) and characteristic frequency (CF), which were defined as the frequency-level combination that elicited the largest response and the frequency that elicited a response at the lowest sound level, respectively. Finally, response bandwidths (BW) were calculated by dividing the CF by the frequency range at 10 dB (Q10) and 30 dB (Q30) above threshold.

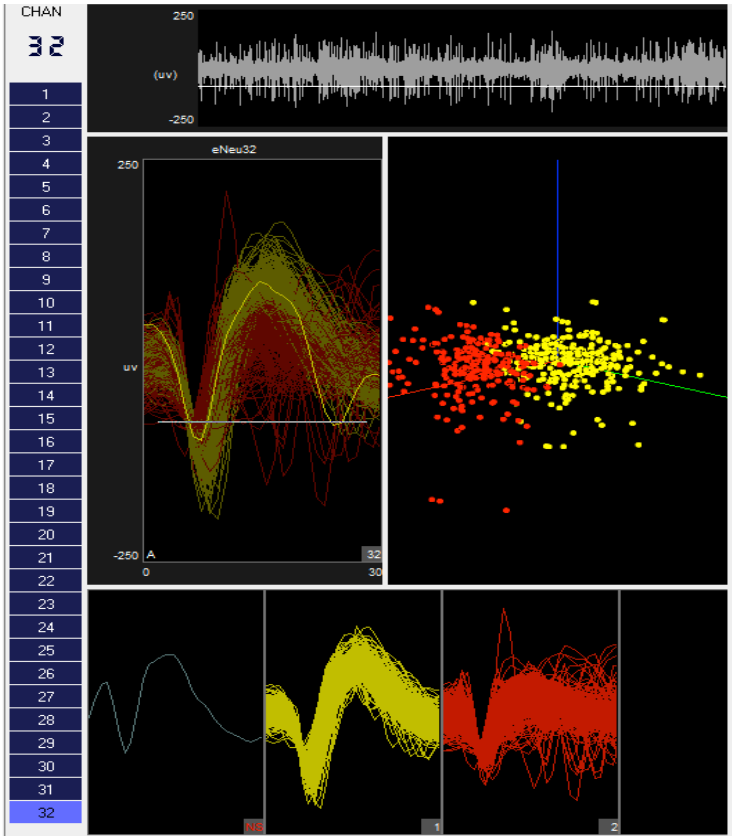


Figure 2-9. Spike sorting method. A screen shot showing the output of the automated Bayesian sorting algorithm for waveforms recorded from channel 32. Top panel shows the raw electrical signal recorded and the threshold set for spike detection. Middle panel shows the shapes of the detected spikes differentiated by colour (left) and their distribution in principal component space (right). Bottom panel shows the waveforms of artefacts (grey) and two putative clusters (yellow, unit 1; red, unit 2).

2.7 Histological Analysis

2.7.1 Tissue Processing

Once no further acoustically driven activity was recorded, animals were perfused transcardially and their brains were extracted for histological analysis. Prior to perfusion, each animal was sedated with Metetomidine Hydrochloride (Domitor, 0.022 mg/kg, IM, Vetoquinol) and then overdosed with Pentobarbital Sodium (Euthatal, 2 ml of 200 mg/ml, IP, Merial Animal Health). Next, the vasculature was washed out with normal saline (0.5 L, 0.9% NaCl), followed with 2L of paraformaldehyde fixative (4%) in 0.1M phosphate buffer ((PB), pH = 7.4). The brain was removed from the cranium and maintained in the same paraformaldehyde fixative for at least 2 days and then placed in a cryoprotective sucrose solution (30% sucrose in 0.1M PB) for a further 3 days. Photographs were taken of the hemispheres to relate the array imprint to the sulcal pattern. The brain was sectioned at 50-µm thickness in the coronal plane using a freezing microtome (Leica SM2000 R sliding microtome, Leica Biosystems, Milton Keynes, UK). Sections were collected in two series, which were stained for Nissl substance and NeuN protein, respectively.

2.7.2 Nissl and NeuN Staining

Nissl staining was performed using cresyl violet acetate, a basic dye, which binds to negatively charged nucleic acids to label all cells. Once completely dried, mounted

sections were placed in cresyl violet acetate (0.5%, 61123, Fluka, Buchs, Switzerland) for 10 minutes and then washed with distilled water for 30 seconds. To achieve differentiation, sections were serially immersed for 1 minute each along a crescendo of concentrated Ethanol (ETOH) solutions: ETOH (50%), ETOH (70%), acidic ETOH (1% glacial acetic acid in 95% ETOH) and ETOH (× 2, 100%). Following differentiation, to reduce unspecific background staining, slides were then serially immersed in chloroform (100%) for 20 seconds, ETOH (× 2, 100%) for 2 minutes and Xylene (× 3, 100%) for 3 minutes. Finally, slides were cover-slipped using mounting medium (Entellan, Merck, Darmstadt, Germany).

NeuN immunostaining was performed to visualise neuronal nuclei. All steps were performed under gentle agitation. Sections were washed repeatedly in 0.1 M phosphate buffer saline (PBS) with 0.1% Triton (PBS-Tx) and incubated in the primary antibody (monoclonal mouse antibody to NeuN, MAB377, Millipore, Temecula, CA, USA) for 3 days at 5°C. After washing repeatedly in PBS-Tx, sections were incubated in the secondary antibody (biotinylated antibody to mouse IgG, Vector Laboratories Inc, Peterborough, UK) for 2 hours at room temperature. Sections were next washed repeatedly in PBS-Tx and then incubated in avidin biotin complex for 90 min at room temperature (PK6100 Vectastain Elite ABC kit, Vector Laboratories Inc, Peterborough, UK). Finally, sections were washed repeatedly in PBS and incubated in a solution containing 0.01% 3,3'-diaminobenzine (DAB, Sigma-Aldrich) and 0.003% H_20_2 in 0.1 M PB until the reaction product was visualised. Sections were mounted on gelatinised slides and cover-slipped with non-aqueous mounting medium (Entellan, Merck, Darmstadt, Germany).

2.7.3 Identification of Recording Sites

The cortical field locations of the recording sites were identified functionally by analysing the FRAs of driven units, which were compared to responses recorded in previously described auditory fields of the ferret auditory cortex (Bizley *et al.*, 2005). In addition, they were identified anatomically by analysing photographic records of the implant imprints on the cortical surface to identify the location of the recording sites relative to the sulci and gyri. In addition, coronal sections stained for Nissl substance and NeuN were studied to determine the depth of electrodes and to correct for angular penetrations. Analysis was performed using a microscope (DMR Leica Microsystems, Heerbrugg, Switzerland) fitted with a digital camera (Microfire, Olympus) and a motorised stage. Neurolucida and Neuroexplorer software (Microbright Field Corp., Williston, VT, USA) were used to map and generate 3-dimensional reconstructions of the cortical surface, layers and electrode tracts.

3 MULTISENSORY VERSUS UNISENSORY LOCALISATION BEHAVIOUR

3.1 Introduction

Perception of events in a natural environment typically depends on multiple facets of information derived from different forms of energy (e.g. sound pressure waves and electromagnetic radiation), with each having specific transduction requirements and encoded within different neural circuits. However, to generate a unified percept, these circuits must converge and their information must be integrated. Multisensory integration has been shown to confer a range of behavioural benefits, from enhanced detection to improved speech comprehension (Murray & Wallace, 2012).

The capacity to merge different sensory signals originating from the same region of space is thought to result in faster and more accurate localisation responses. The analysis of these responses can give a valuable insight into how multisensory

signals are integrated in the brain (Stein & Stanford, 2008). Vision and hearing are generally the most important sensory modalities for perceiving distant objects, with the former providing high-resolution spatial information, while the latter can be used to localise objects and events even if they are not visible because, for example, they are outside of the field of view or obstructed by other objects. However, understanding how spatial information across these two sensory modalities is combined is challenging because visual and auditory signals are encoded in different ways within their respective sense organs and brain pathways. Positions in visual space are represented by the locus of activity within topographic projections from the retina, whereas the auditory system is organised tonotopically, which means that the location of a sound source has to be derived from monaural and binaural spatial cues that arise from the geometry of the head and external ears (King *et al.*, 2001).

A range of behavioural tasks have been used to measure the effects of auditory-visual interactions on localisation behaviour. These include saccadic eye movements with the head still (e.g. Frens *et al.*, 1995; Frens & Van Opstal, 1998; Harrington & Peck, 1998; Colonius & Arndt, 2001; Corneil *et al.*, 2002), combined eye and head gaze shifts (Goldring *et al.*, 1996), head movements (Ho *et al.*, 2013), as well as manual responses (Alais & Burr, 2004; Odegaard *et al.*, 2015). However, these studies have produced variable results in terms of whether combined auditory-visual stimuli actually result in faster and more accurate responses than those elicited by the constituent unimodal sensory stimuli alone. This variation, at least in part, reflects differences in stimulus type and complexity and in task requirements.

In this study, we measured the timing, accuracy and precision of two measures of localisation behaviour to assess the ability of ferrets to integrate

spatiotemporally coincident auditory and visual cues. Sound localisation behaviour has been characterised previously in this species by measuring the initial head orienting response following stimulus presentation and the subsequent locomotor response as the animals approach the perceived location of the sound source in order to receive a water reward (Nodal *et al.*, 2008). Although both behaviours are part of the natural orienting response of the animals, they are differentially affected by lesions (Nodal *et al.*, 2010) or reversible deactivation (Nodal *et al.*, 2012) of the auditory cortex, implying that different neural circuits are involved in guiding each behaviour accurately, which may also differ in their multisensory processes.

3.2 Results

To characterise fully the ability of ferrets to localise stimuli of different modalities (visual, auditory, auditory-visual), the accuracy, precision and timing of their stimulus-evoked head orienting movements and approach-to-target responses were recorded. Although each behaviour can be considered as a component of the localisation response, there are fundamental differences between them, both in terms of the way they are affected by different stimulus parameters and the likely neural circuitry involved. The SC has been shown to play a key role in the generation of head orientation responses (Lomber *et al.*, 2001), whereas previous work in our laboratory has shown that silencing the primary auditory cortical fields impairs approach-to-target localisation (Nodal *et al.*, 2010, 2012). Moreover, the head orienting responses, which were not shaped during the operant training, provide a continuous measure of localisation performance, whereas the conditioned approach-to-target responses represent a categorical measure due to the discrete number of stimulus-response locations.

3.2.1 Localisation Accuracy Varies with the Properties of the Stimulus

On most trials, stereotyped head orienting responses were elicited within 150 ms of stimulus onset. For all stimulus types, the amplitude of these movements co-varied with the eccentricity of the target location, indicating that the animals turned their heads in the direction of the appropriate loudspeaker and/or LED prior to leaving the

central platform (Fig. 3-1). On some trials, the final head bearing matched the target location but, more commonly, it was smaller and therefore under-shot the target.

Head orienting data were analysed by calculating the mutual information values (MI, see Section 2.3.4 for details of method) and fitting a linear regression model to the final head bearing values versus target locations (R^2 range across stimulus durations; visual: 0.17 – 0.35; auditory: 0.36 – 0.77; auditory-visual: 0.53 – 0.71) and the slopes and y-intercepts of these lines compared across stimulus modalities and durations. The above parameters provide a measure of the relative accuracy of the head orienting response, with perfect performance equating to a slope (gain) of 1, a y-intercept (response bias) of 0 and a theoretical maximum MI value of 2.8 bits. For all stimulus types, and in particular the auditory alone condition, increasing stimulus duration tended to result in an increase in the slope of these linear fits and greater MI values (Table 3-1). This was also associated with more precise responses, as indicated by less variation in the final head bearings at longer stimulus durations, particularly for the more eccentric target locations (>30°).

The lowest slopes were observed with visual stimuli, indicating that the animals made smaller head turns compared to the other stimulus conditions. However, if the most eccentric target locations (±90º) are omitted, then the linear regression fits improve for the responses to visual stimuli (R^2 range: 0.20 – 0.46) and the slopes (20 ms: 0.34; 40 ms: 0.41; 100 ms: 0.38; 200 ms: 0.35; 500 ms: 0.55; 1,000 ms: 0.48) now more closely resemble those observed for the auditory and auditory-visual condition. This suggests a similar involvement of head movements in the orienting behaviour of ferrets for all three stimulus types within the frontal

quadrant, corresponding to the 90º wide binocular field of view in this species (Garipis & Hoffmann, 2003). The slopes of the regression fits were very similar for the auditory and auditory-visual stimuli, implying that the head orienting response is driven mainly by auditory stimuli, at least for the most eccentric locations.

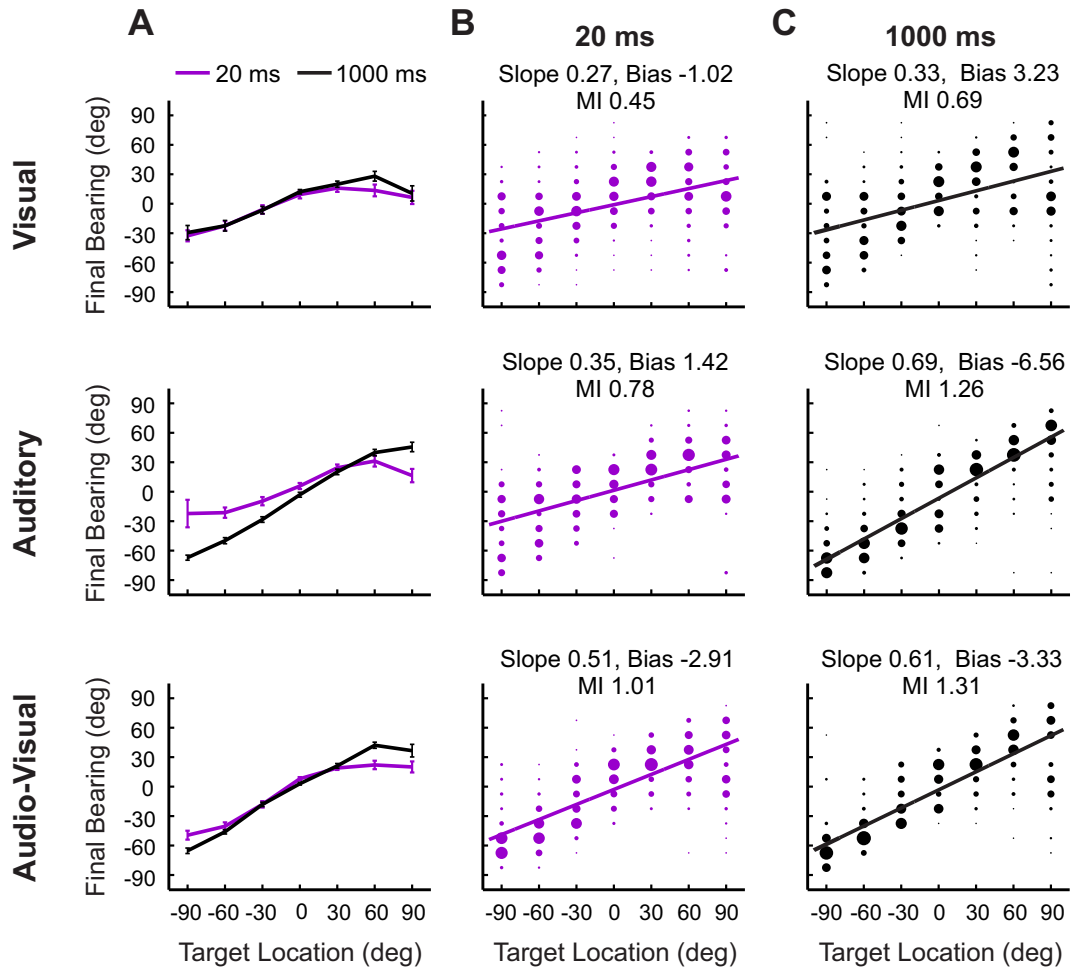


Figure 3-1. Effects of stimulus modality, duration and location on final head bearings. Distribution of final head bearings for each target location across different modalities (top row visual, middle row auditory and bottom row combined auditory-visual) and durations (20 and 1000 ms). (A) Mean final head bearings (± SEM) across each target location for 20 (magenta) and 1000 ms (black) stimuli. (B, C) Confusion matrices for these data illustrating the distribution of final head bearings (bin size 7.5°) as a function of target location. For each stimulus location, the size of the dots is proportional to the probability of responses of different amplitudes. A regression line was fitted (magenta and black lines) to each confusion matrix, using the linear least squares method, and the corresponding slope (gain) and y-intercept (bias) are shown along with the mutual information (MI) values above the panels (see Table 3-1 for a complete set of values). Perfect performance equates to a gain of one, a response bias of zero and a mutual information of 2.8 bits. In general, final head bearings vary less with target location for visual stimuli than for auditory or auditory-visual stimuli.

	Head Orienting			Approach-to-Target		
ms	V	A	AV	V	A	AV
20	0.27 (-1.02) 0.45	0.35(1.42) 0.78	0.51(-2.91) 1.01	0.54 (-7.63) 0.47	0.85 (4.11) 1.36	0.88 (0.86) 1.51
40	0.29 (-2.86) 0.49	0.49 (0.40) 0.97	0.51 (-4.53) 0.98	0.45 (-12.4) 0.50	0.82 (0.82) 1.32	0.85 (2.49) 1.54
100	0.28 (0.28) 0.45	0.47 (-2.30) 0.87	0.49 (-3.00) 0.91	0.50 (-11.1) 0.59	0.84 (-0.80) 1.33	0.88 (-0.72) 1.67
200	0.26 (-10.1) 0.52	0.53 (-2.04) 1.02	0.47 (-2.24) 0.95	0.42 (-10.2) 0.60	0.82 (-0.83) 1.43	0.90 (-1.74) 1.88
500	0.39 (-3.15) 0.69	0.53 (-0.12) 1.15	0.62 (-5.55) 1.35	0.70 (-3.40) 1.20	0.87 (-0.80) 1.78	0.93 (-2.40) 2.23
1000	0.33 (3.23) 0.69	0.69 (-6.56) 1.26	0.61 (-3.33) 1.31	0.73 (-4.07) 1.57	0.93 (-1.09) 2.05	0.96 (-0.23) 2.39

Table 3-1. Linear model parameters and mutual information values for head orienting reactions and approach-to-target responses. Regression lines were fitted to plots of final head bearing values and approach-to-target response locations versus target locations. The slopes and y-intercepts (in parentheses) of these regression lines are shown on the top line along with the mutual information in bits between response and target locations on the bottom line for each stimulus duration (rows, 20 – 1000 ms) and type (columns, visual (V), auditory (A), auditory-visual (AV)).

In order to explore the effect of stimulus modality on head orienting responses, we calculated the final head bearing errors, i.e. the difference between the final head bearing and the target location (Fig. 3-2). Final head bearing errors were standardised so that an undershoot resulted in a negative value and an overshoot in a positive value, regardless of whether the stimulus was presented from either the left or the right of the midline. Overall, head-orienting accuracy was greatest for the auditory-visual condition, as demonstrated by the smallest error magnitude (visual = -41.2 ± 3.2°; auditory = -27.9 ± 2.7°, auditory-visual = -27.6 ± 2.3°). Across all

stimulus modalities, smaller error magnitudes were observed at longer stimulus durations (Fig. 3-2A) and more central locations (Fig. 3-2B).

A linear mixed effects analysis was performed to examine the effects of stimulus modality, duration and location on final head bearing errors (Fig. 3-3). The model showed that final head bearing errors were significantly smaller for auditory-visual than visual stimuli (see appendix 3.4.1.1 for significance values) for the majority of stimulus duration-location combinations, with the exception of -30° at 1000 ms, 30° at 500 – 1000 ms and 90° at 20 ms. In contrast, final head bearing errors were not significantly smaller for auditory-visual than for auditory stimuli (see appendix 3.4.1.1 for significance values) for the majority of stimulus duration-location combinations, with the exception of -90° at 20 - 200 ms, -60° at 20 – 100 ms and -30° at 20 – 40 ms. This finding suggests that, under multisensory conditions, the head-orienting response may be driven primarily by the auditory modality.

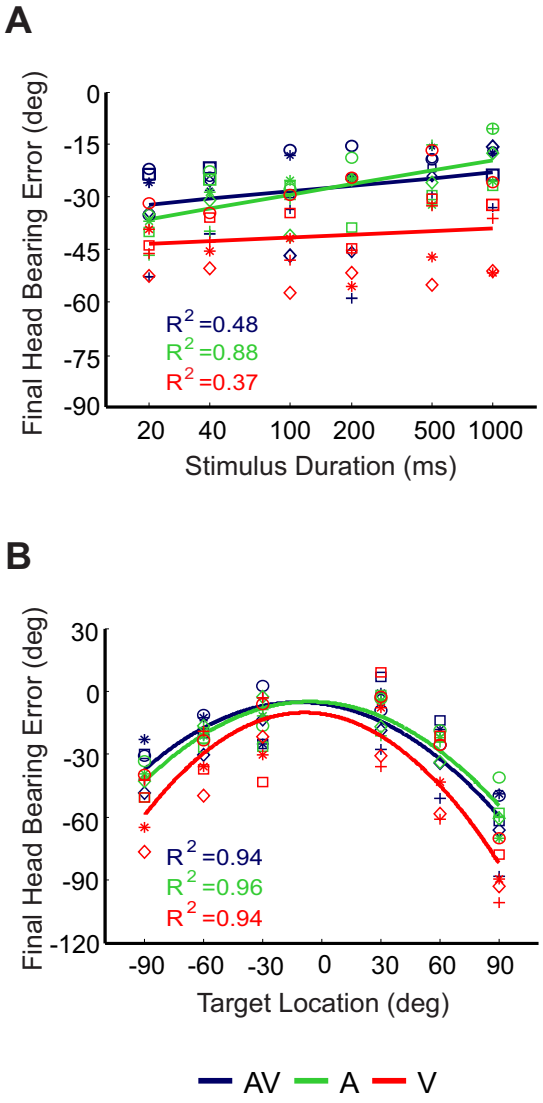


Figure 3-2. Comparison of head orienting accuracy to unisensory and multisensory targets. Note that final head bearing errors were standardised so that an undershoot relative to the target location resulted in a negative value and an overshoot in a positive value, regardless of whether the stimulus was presented from either the left or the right of the midline. Head orienting errors are shown as a function of stimulus duration (A), plotted on a log scale, and stimulus location (B) for auditory-visual (AV, blue), auditory (A, green) and visual (V, red) targets. Mean final head bearing errors were fitted with a regression line (A) or a polynomial function (B). The relevant coefficient of determination (R^2) value is shown for each dataset. Data from individual animals are represented by different symbols. Note that head orienting accuracy was similar for the auditory-visual and auditory targets across all stimulus durations and locations.

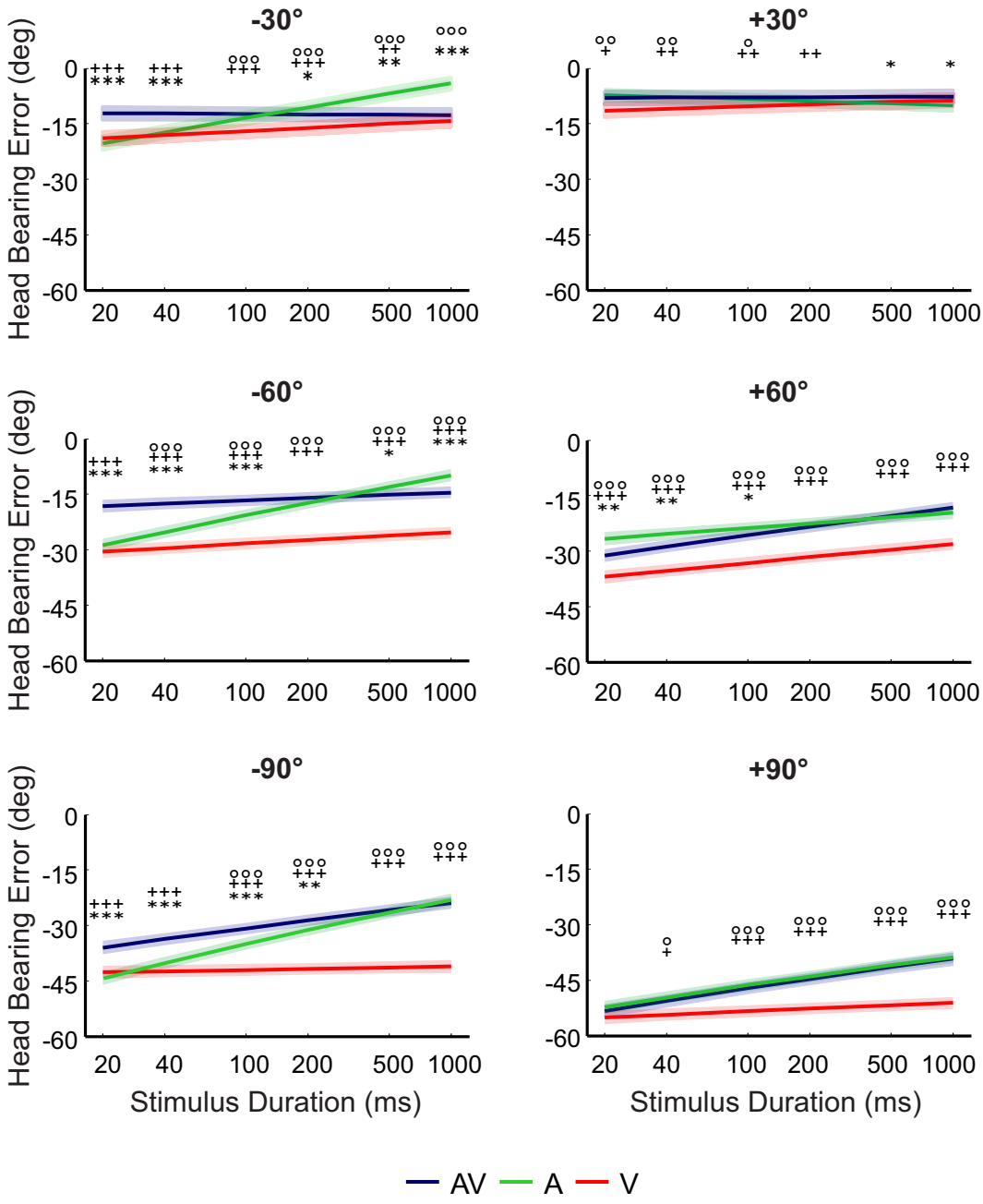


Figure 3-3. Results of the linear mixed effects model fitted to the final head bearing error data. Final head bearing errors (± 95% CI) predicted by the model are plotted as a function of stimulus duration (on a log scale) for auditory-visual (AV, blue), auditory (A, green) and visual (V, red) targets. Results for each target location are shown in separate panels. Significant differences between stimulus conditions are indicated by the symbols (asterisks: AV vs. A; crosses: AV vs. V; circles: A vs. V and *P*-value by their number, one: $P < 0.05$; two: $P < 0.01$ and three: $P < 0.001$).

Approach-to-target localisation accuracy, measured using probability correct scores, improved across all sensory conditions as stimulus duration was increased (Figs. 3-4 and 3-5). MI values and the slopes and y-intercepts of fitted regression lines (R^2 range across stimulus durations; visual: 0.17 – 0.54; auditory: 0.68 – 0.89; auditory-visual: 0.75 – 0.94) were calculated to quantify the relationship between target and response location and these values were compared across stimulus modalities and durations (Table 3-1). For each stimulus type, both the MI values and slopes tended to increase with increasing stimulus duration, mirroring the improvement in probability correct scores. As with the head orienting responses, the approach-to-target responses also became more precise at longer stimulus durations, particularly for the more eccentric target locations (>30°; Fig. 3-4). A clear difference was found in approach-to-target localisation accuracy for the different stimulus types (Fig. 3-4 and Table 3-1). For all stimulus durations, the highest probability correct scores were obtained for the auditory-visual targets, followed by the auditory targets and then the visual targets. The same pattern was observed by comparing the slopes of the regression lines and the MI between the response and target locations (Table 3-1).

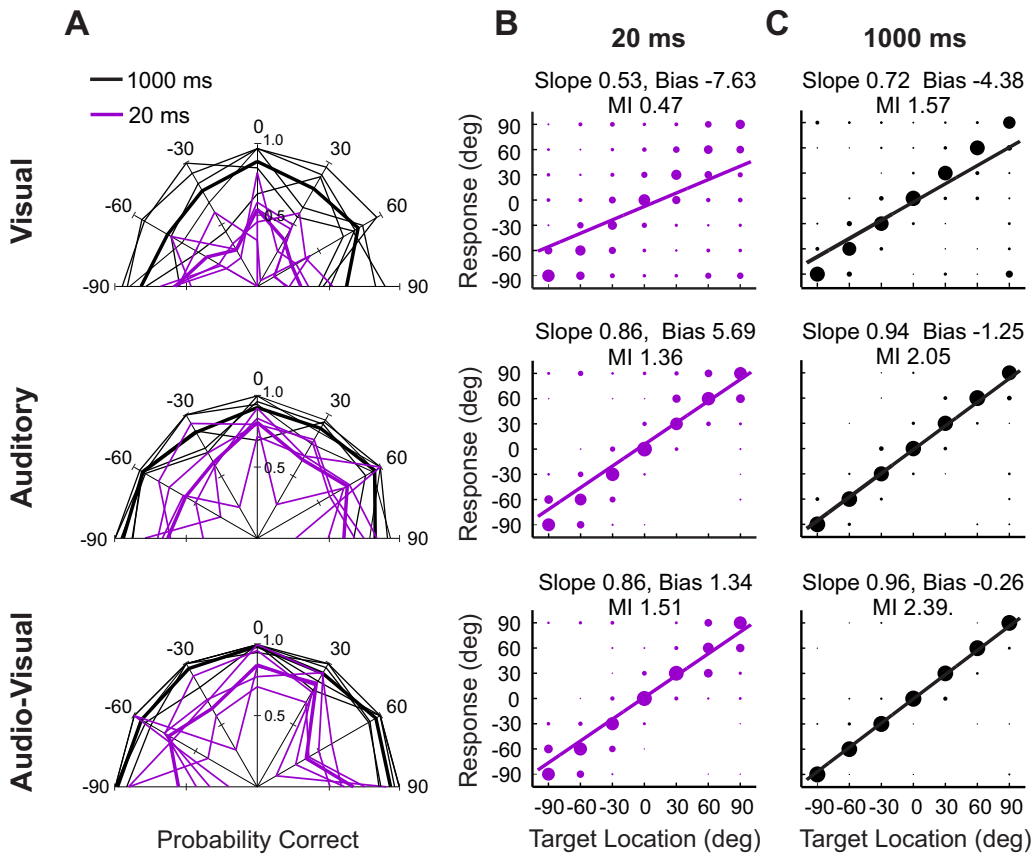


Figure 3-4. Effects of stimulus modality, duration and location on approach-to-target responses. The distribution of approach-to-target responses is shown for each target location across different stimulus modalities (top row visual, middle row auditory and bottom row auditory-visual) and durations (20 and 1000 ms). (A) Approach-to-target probability correct scores at each target location for stimulus durations of 20 ms (magenta lines) and 1000 ms (black lines). The thin lines indicate the performance of individual animals; thick lines represent the overall mean performance. (B, C) Confusion matrices for these data illustrating the distribution of approach-to-target responses. The size of the dots indicates for each stimulus location the proportion of responses made to different target locations. A regression line was fitted (magenta and black) to each confusion matrix, using the linear least squares method, and the corresponding slope (gain) and y-intercept (bias) are shown along with the mutual information (MI) values above the panels (see Table 3-1 for a complete set of values).

To quantify the magnitude of the multisensory enhancement of approach-to-target accuracy, multisensory gains were calculated (Fig. 3-5A,B top panels). Overall, gains were positive when compared against either of the unisensory stimulus conditions and were larger relative to the visual than to the auditory stimuli (auditory-visual vs. visual = 43.17 %; auditory-visual vs. auditory = 9.96 %). Multisensory gains relative to the auditory stimulus showed no clear correlation with stimulus duration or location, indicating that the improvement in performance was present across the full range of stimulus durations tested (Fig. 3-5A) and throughout the frontal hemifield (Fig. 3-5B). In contrast, multisensory gains relative to the visual stimulus were relatively uniform at shorter stimulus durations (≤ 200 ms) and declined at longer durations (500 – 1000 ms) (Fig. 3-5A), whereas no clear differences in gain for multisensory versus visual stimuli were apparent for the different locations within the frontal hemifield (Fig. 3-5B).

To investigate the significance of this multisensory enhancement effect across different stimulus durations and locations, a generalised linear mixed effects model for Bernoulli-distributed responses was fitted to the approach-to-target response correctness data (Fig. 3-6). The results show that animals were significantly more likely to make a correct approach-to-target response to auditory-visual than to visual stimuli for all but one (-90° at 20 ms) stimulus combinations (see Fig. 3-6 and appendix 3.4.1.2 for significance values). Similarly, they were significantly more likely to respond correctly to auditory-visual than to auditory targets across the majority of stimulus combinations, with the exception of -90° at 20 - 40 ms, 0° at 20 ms and 60° at 20 - 500 ms (see Fig. 3-6 and appendix 3.4.1.2 for significance values).

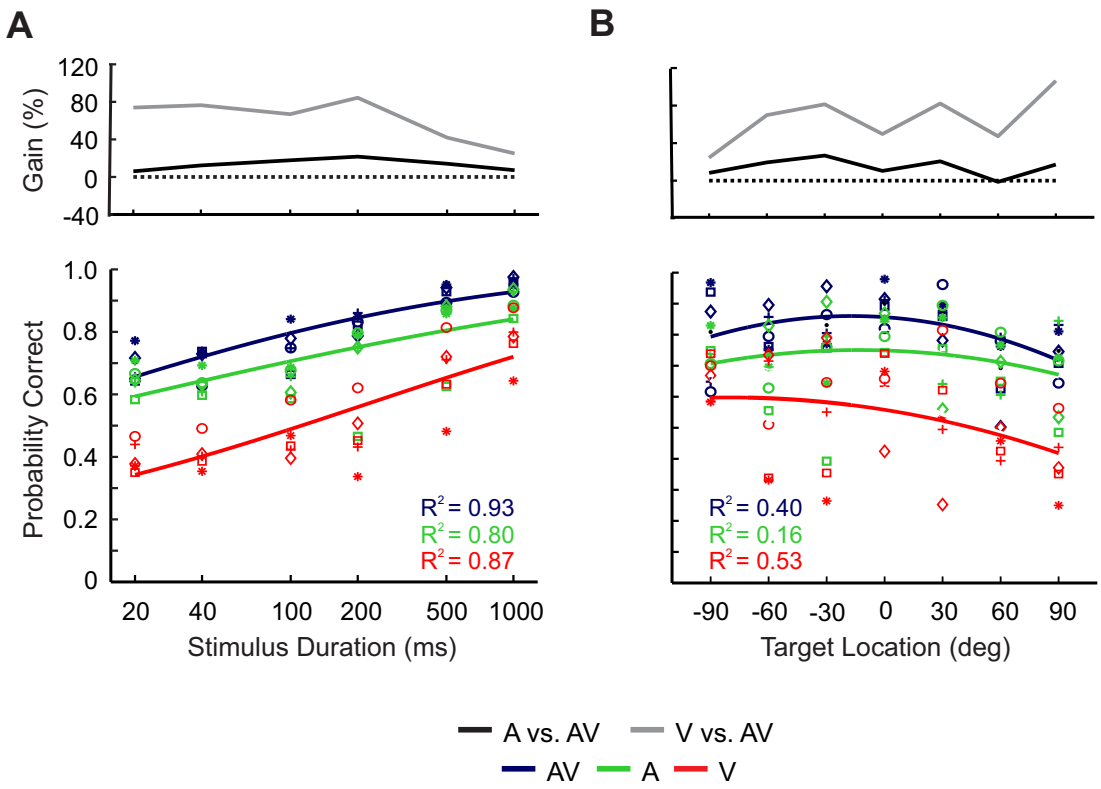


Figure 3-5. Comparison of approach-to-target localisation accuracy for unisensory and multisensory targets. Approach-to-target probability correct scores are shown as a function of stimulus duration (A, bottom panel), plotted on a log scale, and stimulus location (B, bottom panel) for auditory-visual (AV, blue), auditory (A, green) and visual (V, red) stimulus targets. Mean probability correct scores were fitted with a cumulative normal distribution or a polynomial function. The relevant coefficient of determination (R^2) value is indicated for each dataset. Individual animals are represented by the different symbols. In contrast to the head orienting responses, approach-to-target localisation accuracy was greater for the auditory-visual targets than for either the auditory or visual targets across all stimulus durations and locations. The multisensory percentage gains for the approach-to-target responses are shown as a function of stimulus duration (A, top panel) and stimulus location (B, top panel) for auditory versus auditory-visual (black) and visual versus auditory-visual (grey) targets. A positive gain indicates that the mean performance for the multisensory condition is superior to that for the unisensory condition.

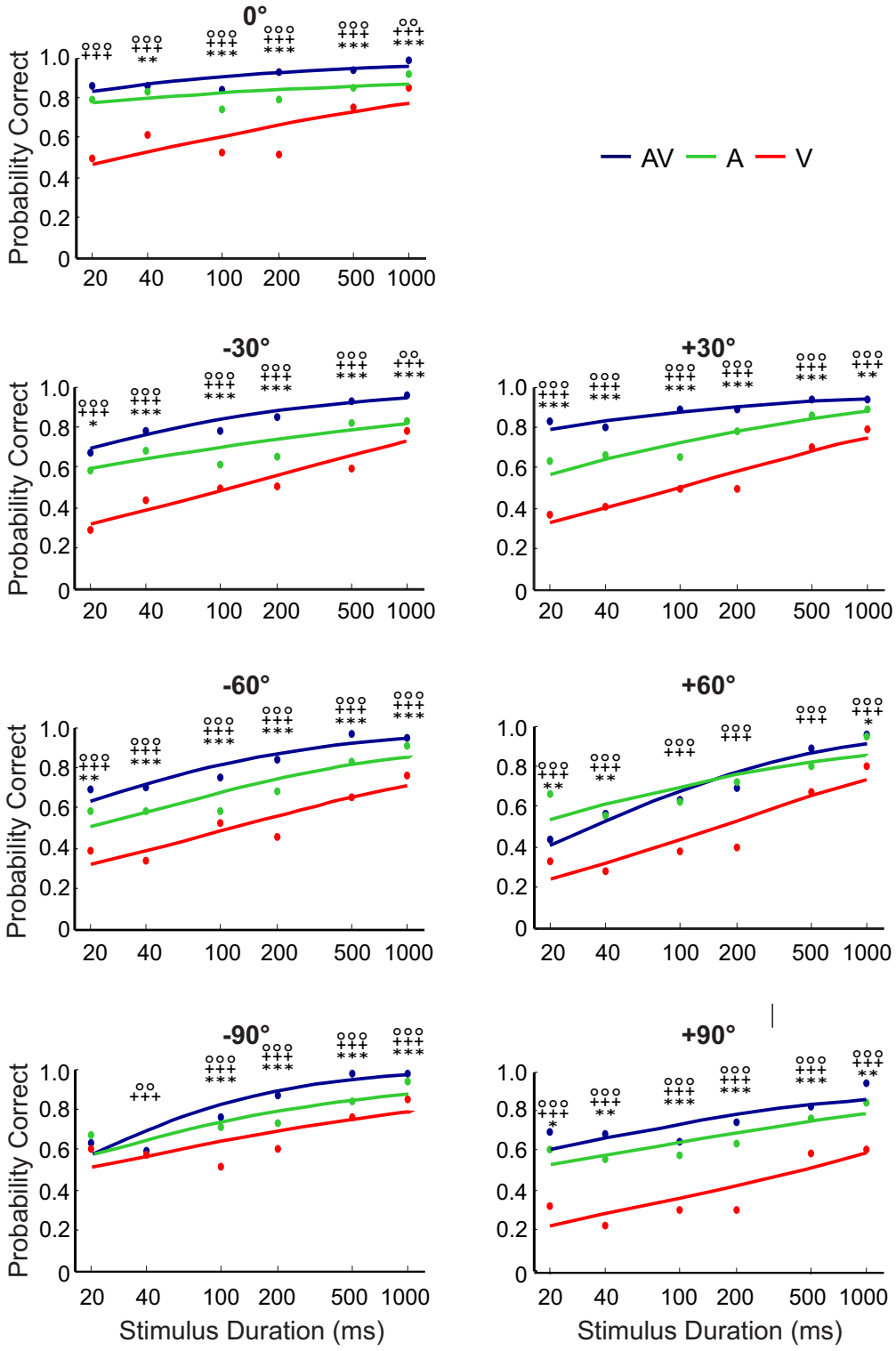


Figure 3-6. Results of the generalised linear mixed effects model fitted to the approach-to-target accuracy data. Modelled probabilities of a correct approach-to-target response plotted as a function of stimulus duration (on a log scale) for auditory-visual (AV, blue), auditory (A, green) and visual (V, red) targets. Results for each target location are shown in separate panels, together with the actual probabilities for

comparison (filled circles). Significant differences between stimulus conditions are indicated by the symbols (asterisks: AV vs. A; crosses: AV vs. V; circles: A vs. V) and the *P*-value by their number (one: $P < 0.05$; two: $P < 0.01$ and three: $P < 0.001$).

3.2.2 Reaction and Response Times

In addition to measuring localisation accuracy, head orienting reaction times and approach-to-target response times were analysed to assess whether ferrets demonstrated a redundant signal effect (RSE) (Todd, 1912), according to which they should react and respond more rapidly to auditory-visual than to unisensory stimuli.

A correlation between response times and the correctness of approach-to-target responses has previously been described in ferrets using a 360° auditory localisation task, whereby correct responses were shown to be faster than incorrect responses (Nodal *et al.*, 2008). To investigate whether this effect exists across the frontal hemifield for reaction and response times from the present task, separate generalised linear models for Bernoulli-distributed responses were fitted to the approach-to-target correctness data (Fig. 3-7). The results show that, across stimulus modalities and durations, head orienting reaction times (auditory-visual: $t_{16769} = -4.99$, $P < 0.001$; auditory: $t_{16769} = -6.59$, $P < 0.001$; visual: $t_{16769} = -7.07$, $P < 0.001$) and approach-to-target response times (auditory-visual: $t_{14855} = -7.75$, $P < 0.001$; auditory: $t_{14855} = -9.70$, $P < 0.001$; visual: $t_{14855} = -9.73$, $P < 0.001$) were significantly negatively correlated with the probability of a correct approach-to-target response. On some trials, this effect might occur simply because the animal fails to register the stimulus and goes on to approach a random reward spout. However, on the majority

of trials in which an incorrect response was made the error magnitude was 30°, i.e. the animal approached the adjacent reward spout, which suggests a lack of precision rather than a failure to register the stimulus. Instead it is possible, that when uncertain of the correct response, animals attempt to alter their performance strategy by sacrificing response speed for a perceived improvement in performance, a strategy known as the speed-accuracy trade-off (Fitts, 1954; Leach *et al.*, 2013).

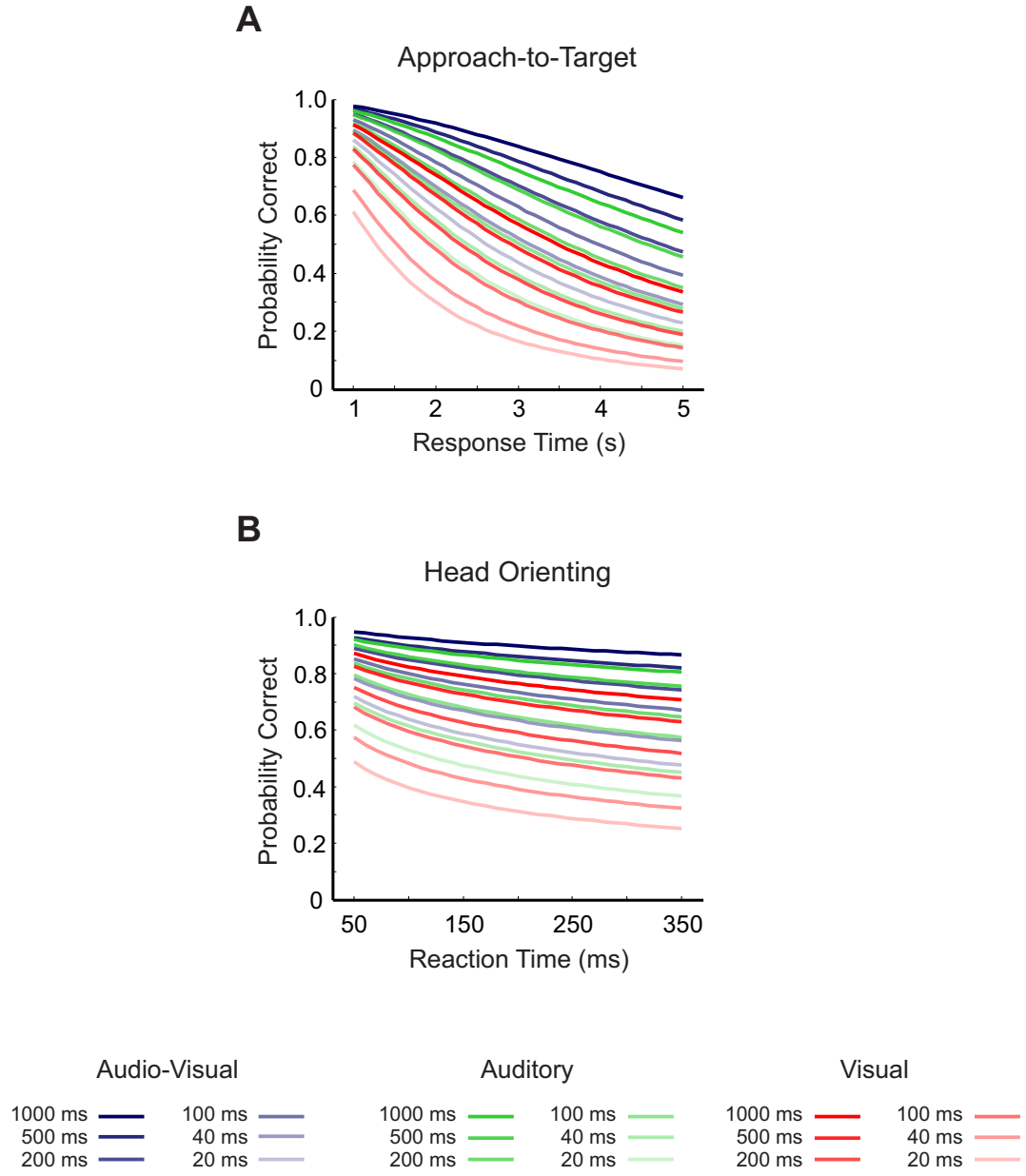


Figure 3-7. Approach-to-target localisation accuracy versus reaction and response times. Modelled probability of a correct approach-to-target response plotted against head orienting reaction (A) and approach-to-target response (B) times for auditory-visual (AV, blue), auditory (A, green) and visual (V, red) targets. In general, for all stimulus types and across stimulus durations, incorrect responses were associated with longer reaction and response times than correct responses.

In view of this result, together with the varying proportion of correct trials observed between stimulus types, incorrect trials were omitted from the following analysis. Furthermore, to prevent any bias due to a possible lack of attention, which might have resulted in abnormally slow responses, reaction times > 250 ms and response times > 5 s were also omitted from the analysis. In total, 5.4% of reaction times and 1.3% of response times were excluded because of these criteria. Overall, reaction times (auditory-visual: 78.82 ± 0.55 ms, auditory: 87.05 ± 0.63 ms, visual: 142.10 ± 1.01 ms) and response times (auditory-visual: 1.74 ± 0.01 s, auditory: 1.99 ± 0.01 s, visual: 1.97 ± 0.01 s) were shorter for the auditory-visual stimuli than for either of the unisensory stimuli. For all animals, increasing the stimulus duration had no noticeable effect on reaction and response times (Fig. 3-8A), however, varying the stimulus location had a noticeable effect on reaction and response times (Fig. 3-8B), whereby the shortest times tended to occur for the central locations and -90°.

To investigate the significance of this RSE across the different stimulus locations, separate linear mixed effects models were fitted to the reaction and response time data (Fig. 3-9). The results show that ferrets took significantly less time to reach auditory-visual targets than either visual or auditory targets (see Fig. 3-9B and appendix 3.4.1.3 for significance values). In contrast, although head orienting reaction times were significantly shorter for auditory-visual stimuli than for visual stimuli, with the exception of the -30° stimulus location, a multisensory advantage was not seen relative to the auditory stimuli across the majority of stimulus locations with the exception of -60° and -90° (see Fig. 3-9A and appendix 3.4.1.4 for significance values). These differences are supported by the finding that reaction times were significantly shorter for auditory than for visual stimuli at the majority of

stimulus locations, whereas for approach-to-target response times this was not the case. For both reaction (t_{10212} = -1.22, P = 0.22) and response (t_{15832} = 0.360, P = 0.72) times, no significant effect of stimulus duration was identified. The multisensory influence on the speed of the responses made on correct trials was quantified by calculating the multisensory gain. Overall, gains were positive for both reaction times (auditory-visual vs. auditory = 7.24 %; auditory-visual vs. visual = 41.19 %) and response times (auditory-visual vs. auditory = 11.98 %; auditory-visual vs. visual = 12.39 %) when compared with either form of unisensory stimulation.

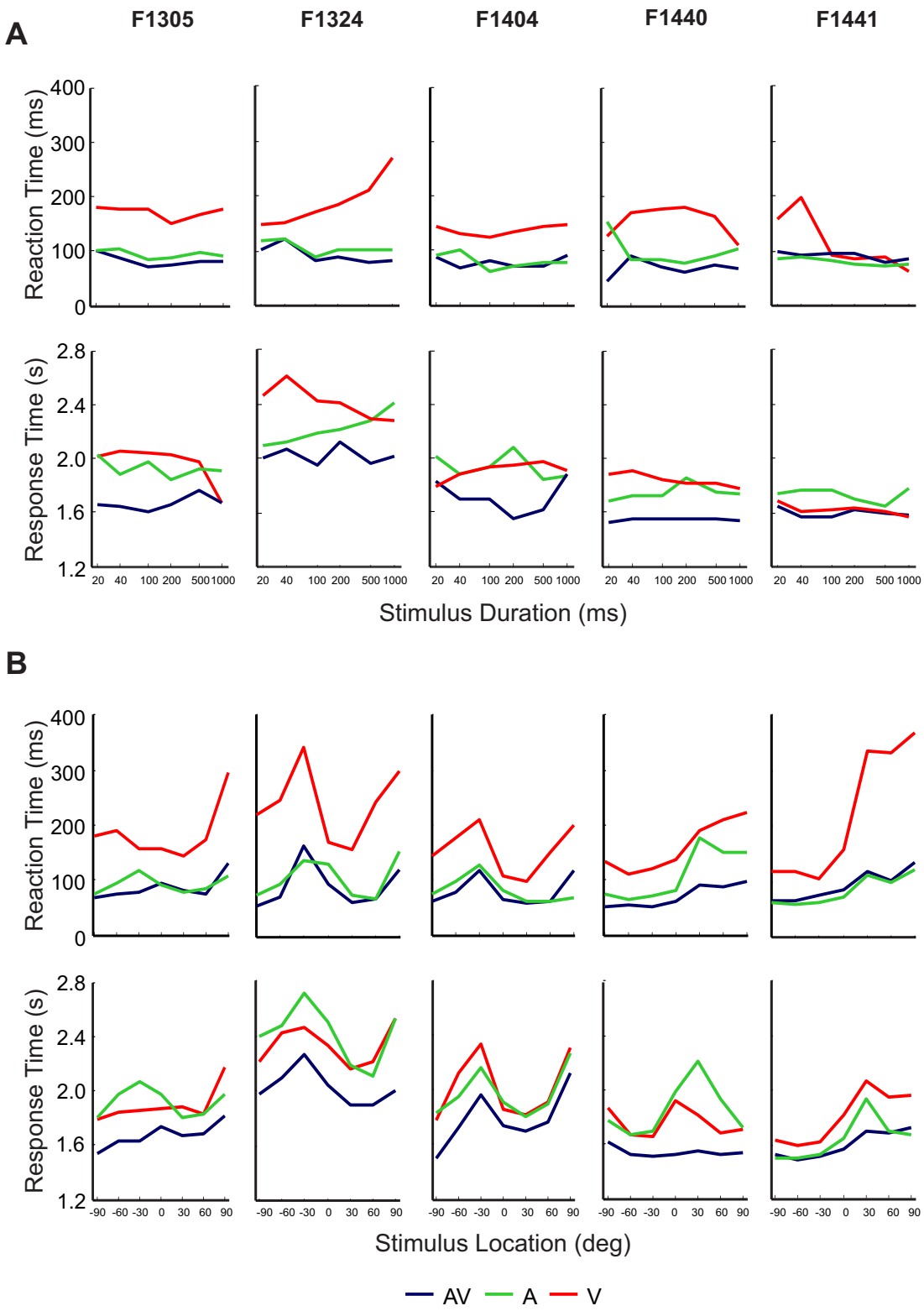


Figure 3-8. Effects of stimulus modality, duration and location on head orienting reaction times and approach-to-target response times. Individual reaction and response times plotted as a function of stimulus duration (A, on a log scale) and location (B) for auditory-visual (AV, blue), auditory (A, green) and visual (V, red) targets.

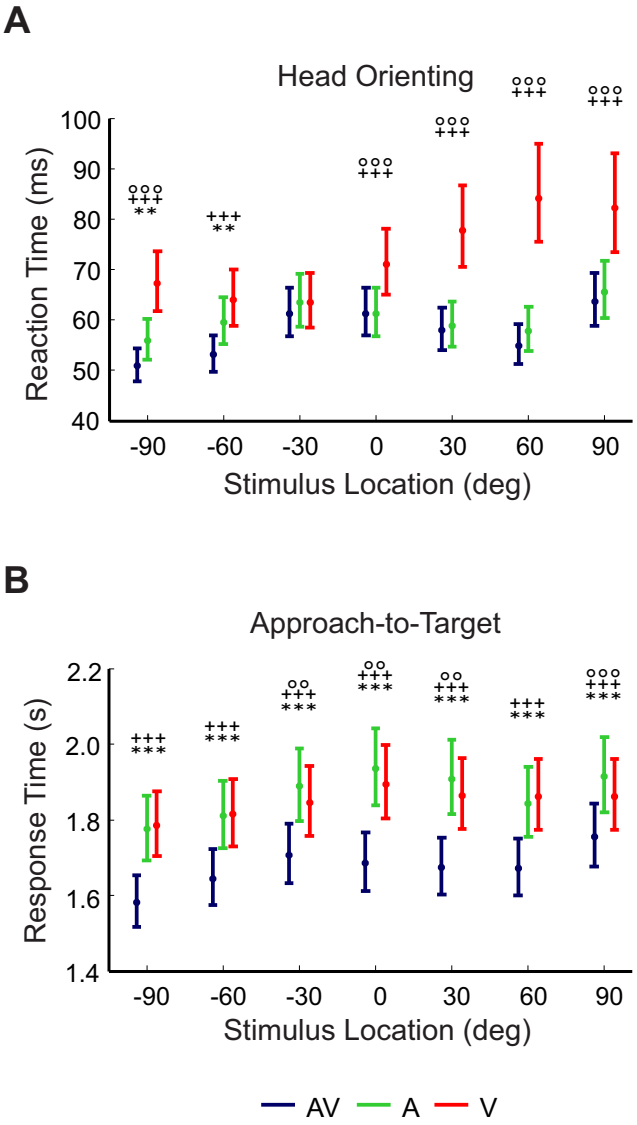


Figure 3-9. Comparison of head orienting reaction and approach-to-target response times for unisensory and multisensory targets. Model predicted (± 95% CI) head orienting reaction times (A) and approach-to-target response times (B) are plotted against stimulus location for auditory-visual (AV, blue), auditory (A, green) and visual (V, red) targets. Note that reaction and response times were shorter across all stimulus locations for auditory-visual targets than for either auditory or visual targets. Significant differences between stimuli are indicated by the symbols (asterisks: AV vs. A; crosses: AV vs. V; circles: A vs. V) and the *P*-value by their number (one: $P < 0.05$; two: $P < 0.01$ and three: $P < 0.001$).

Finally, data were subjected to race model inequality analysis (Raab, 1962) to investigate the basis for the observed RSE (Fig. 3-10). It is possible to derive the maximum multisensory facilitation of reaction and response times that can be expected given a race scenario by taking the sum of the auditory and visual reaction (or response) time cumulative distribution functions (CDF) (see Section 2.3.4). The model is violated if, at any time, the observed auditory-visual CDF exceeds the modelled CDF, which by default implies the existence of a signal integration mechanism (Miller, 1982). Analysis of approach-to-target response times showed that the model was consistently violated over the first seven deciles (Fig. 3-10B, D), suggesting that multisensory integration in the brain could be responsible for the faster responses. As might be expected, significant violations were reliably observed at the lower deciles (the 1st to the 5th), since summation of auditory and visual inputs would most likely result in faster times by reaching the response threshold earlier than either input in isolation. On the other hand, analysis of reaction times showed no violation of the model (Fig. 3-10A,C), implying that the multisensory facilitation of reaction times could be explained in terms of probability summation alone.

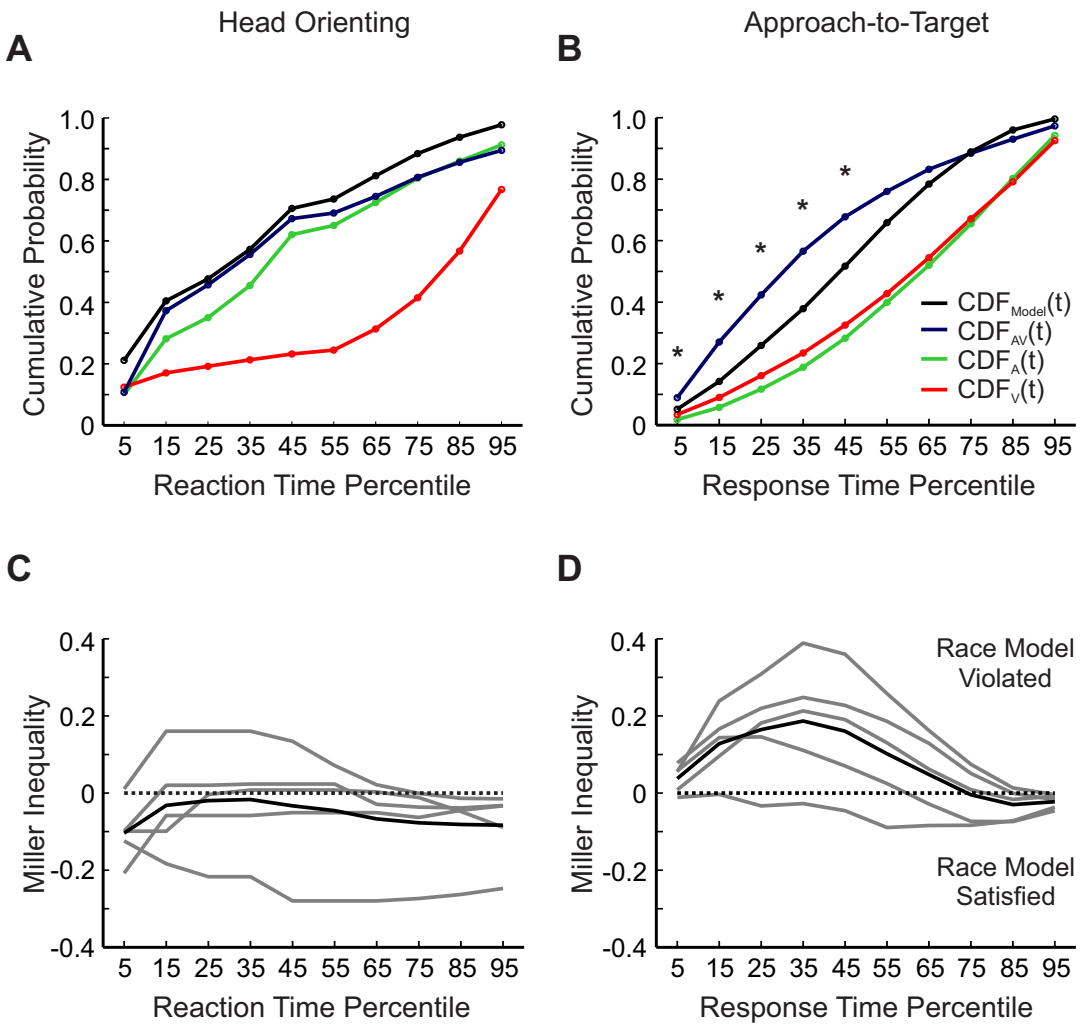


Figure 3-10. Race model analysis. Group aggregated cumulative probability distribution functions (CDF) of head orienting reaction times (A) and approach-to-target response times (B), displayed for auditory-visual (AV, blue), auditory (A, green) and visual (V, red) targets and for the race model (black). In the case of the approach-to-target data, the model can be rejected since the auditory-visual CDF (blue line) exceeds the modelled CDF (black line), with significant differences across response time percentiles marked by asterisks (*, $P < 0.05$). (C, D) Results of Miller's inequality ($CP_{(t)AV} - CP_{(t)Model}$) test of the race model (see Methods for details) for head orienting reaction times (C) and approach-to-target response time (D): positive values indicate violation of the model and negative values its satisfaction. Individual animals are shown by the grey lines; group aggregated values are shown by the black lines.

3.3 Discussion

The capacity to merge sensory signals from different modalities has been shown to confer a range of behavioural advantages. A variety of behavioural tasks have been used to measure the effects of auditory-visual interactions on localisation behaviour including saccadic eye movements with the head still (e.g. Frens *et al.*, 1995; Frens & Van Opstal, 1998; Harrington & Peck, 1998; Colonius & Arndt, 2001; Corneil *et al.*, 2002), combined eye and head gaze shifts (Goldring *et al.*, 1996), head movements (Ho *et al.*, 2013), as well as manual responses (Alais & Burr, 2004; Odegaard *et al.*, 2015). However, these studies have produced variable results in terms of whether combined auditory-visual stimuli actually result in faster and more accurate responses than those elicited by the constituent unimodal sensory stimuli.

In this chapter, the ability of ferrets to localise unisensory and multisensory stimuli of different durations was characterised by measuring both their head orienting and approach-to-target behaviour. Approach-to-target localisation responses were more accurate and faster to spatially congruent auditory-visual stimuli throughout the frontal hemifield than to either visual or auditory stimuli presented in isolation. Conversely, whilst head orienting responses to auditory-visual stimuli were more accurate and faster than those made to visual targets, this was not the case when they were compared to the responses made to auditory targets, suggesting that these movements were guided principally by sound alone.

For distant objects, vision generally provides higher resolution spatial information than audition (DeValois & DeValois, 1991; Brown & May, 2005).

Indeed, ferrets can resolve drifting sine-wave gratings with a spatial frequency of 0.5 cycles per degree (von Melchner *et al.*, 2000), whereas their minimum audible angles are ~10° at the midline (Kavanagh & Kelly, 1987b; King & Parsons, 1999). Nevertheless, we observed that ferrets were able to localise single presentations of broadband noise more accurately than light flashes of equivalent duration from LEDs. This was observed in both the head orienting and approach-to-target data and across all stimulus durations and locations and, therefore, cannot be attributed to the most eccentric visual stimuli falling outside the field of view. This is unlikely to be due to the way they were trained, since all animals were trained initially using auditory-visual stimuli and were not tested until stable criterion levels of performance were achieved with each stimulus type. Although a guiding role for visual cues in the plasticity of auditory spatial processing has been demonstrated during development (King *et al.*, 1988) in this species, it is possible that the orienting behaviour of ferrets naturally relies more on their hearing than their sight. This is consistent with our finding that, for the stationary stimuli used in this study, the latency and accuracy of their head turns to spatially congruent auditory-visual stimuli were not significantly different from those made to sound alone.

Head movements following stimulus presentation were used as a metric for assessing the initial orienting response of the ferrets. The head movements directed toward visual and auditory targets had a comparable range of amplitudes and latencies to those described in cats (Ruhland *et al.*, 2013), with reaction times ~50 ms shorter when the animals localised a sound source. We found that the relationship between final head bearing and target location was similar for the different stimulus types, with the notable exception that head movements to the most peripheral visual

stimuli were smaller than those made to auditory or paired auditory-visual targets. Again, this is unlikely to be due to the animals not seeing the most eccentric visual targets (which extended to ±90°), since the visual field in ferrets extends to at least 110° (King & Hutchings, 1987).

This modality-dependent difference in head orienting could reflect a smaller contribution of head movements (as opposed to eye-in-head movements) to the gaze shifts made to visual than to auditory stimuli, as has been reported in cats (Ruhland *et al.*, 2013). However, we think this is unlikely since the ferret eye contains a prominent visual streak and a much lower variation in retinal ganglion cell density across the retina (Stone, 1965; Henderson, 1985), raising the possibility that this species relies less on eye movements than other carnivores. Indeed, the lateral rectus muscle, which is responsible for abduction of the eye, has a much slower twitch contraction time and is innervated by relatively few abducens nucleus motor neurons as compared to cats (Bishop *et al.*, 1985). Moreover, the approach-to-target responses showed a comparable dependence on stimulus location and duration and were also consistently more accurate for auditory than for visual stimuli. Although direct measurements of eye position in freely-moving animals will be required to verify this, it seems reasonable to assume that head movements alone provide an appropriate measure of localisation performance. Whilst the ferrets could localise 20-ms light flashes from an LED positioned 70 cm away at well above chance levels, the noise bursts presumably provided more salient cues at all durations tested.

In humans, combined eye-head gaze shifts tend to have significantly shorter latencies than the shortest unisensory reaction time across a range of eccentricities

(Goldring *et al.*, 1996), whereas we observed no difference for most target locations between the head movements made to auditory and auditory-visual targets. Consequently, the multisensory gain was small and could be accounted for by the race model, implying that the orienting movements are triggered by the brain's response to the auditory stimulus rather than the later response to the visual stimulus. However, the relative metrics of eye and head movements can vary with stimulus modality and target location in complex ways (Goldring *et al.*, 1996; Ruhland *et al.*, 2013), so we cannot rule out the possibility that a clearer multisensory advantage would have been observed had eye-head gaze shifts been measured.

In contrast, measurements of approach-to-target behaviour provided robust evidence for neural integration of auditory and visual inputs. Combining auditory and visual stimuli at the same location has been shown to improve localisation accuracy in humans (Alais & Burr, 2004; Odegaard *et al.*, 2015) and cats (Stein *et al.*, 1989a; Gingras *et al.*, 2009). Although these studies did not report response times, we found that ferrets not only localised auditory-visual stimuli more accurately than either modality by itself, but they also made faster responses. Indeed, the measured response times were shorter than those predicted by the race model, indicating an effect of multisensory convergence on the neural processing of spatial information. In the present study, all procedural training was carried out using auditory-visual stimuli to avoid introducing a training bias towards either sensory modality. However, the results suggest that animals relied primarily on the auditory stimuli for their behaviour. Therefore, it is possible that a stronger multisensory effect might have been observed if the animals had been trained using only one modality and then tested using both modalities. However, we think this is unlikely

since the order of the testing blocks was reversed for each animal (AV followed by A versus A followed by AV), yet the multisensory enhancement effects observed across animals were relatively constant.

A number of brain regions have been implicated in the integration of auditory and visual cues. Spatial information from different sensory modalities is combined in the superior colliculus (SC), which plays a key role in the generation of eye-head gaze shifts (Freedman & Sparks, 1997; Bell *et al.*, 2005; Gandhi & Katnani, 2011) and orientation responses (Lomber *et al.*, 2001). For example, the ability of cats to orient towards and approach visual and auditory targets in the frontal hemifield is impaired following reversible inactivation of the superficial and intermediate layers of the SC, respectively (Lomber *et al.*, 2001), while excitotoxic lesions of the latter result in a more persistent loss of the modulatory effect of one stimulus modality on another in this task (Burnett *et al.*, 2004). The ferret SC conforms to the general vertebrate plan whereby visual and auditory representations are organised topographically to form overlapping maps of sensory space (King & Hutchings, 1987), with neurons in the deeper layers of this midbrain nucleus often responsive to stimuli in more than one sensory modality (King & Schnupp, 2000; Meredith *et al.*, 2000). More extensive studies in other species have shown that the activity of these multisensory neurons can be altered when different stimuli are combined, with the strongest response enhancement occurring for visual and auditory stimuli that are roughly coincident in time and space (King & Palmer, 1985; Meredith & Stein, 1986a; Meredith *et al.*, 1987; Wallace *et al.*, 1996). The largest multisensory responses have been shown to occur when the visual stimulus is presented before the auditory stimulus. This offset correlates with the temporal disparity between the

unisensory response latencies, which can be 44 - 60 ms longer for visual stimuli (Miller, 1982; Wallace *et al.*, 1998). This temporal dependence could explain the lack of race model violation for the reaction times particularly at shorter stimulus durations, given that the duration of stimuli are potentially shorter than their latency difference.

However, the SC does not act alone in mediating crossmodal influences on localisation behaviour, since these effects are also reduced in cats following inactivation of the anterior ectosylvian sulcus (AES), a multisensory cortical region that innervates the SC (Wilkinson *et al.*, 1996). It is also possible that other cortical regions are involved since there is now extensive evidence for multisensory convergence and interactions in primary and non-primary sensory areas (e.g. Falchier *et al.*, 2002; Rockland & Ojima, 2003; Bizley *et al.*, 2007; Lakatos *et al.*, 2007; Hall & Lomber, 2008; Kayser *et al.*, 2010; Henschke *et al.*, 2015). The auditory cortex is likely to be particularly important in this respect since spatially-coincident visual cues can enhance spatial processing by auditory cortical neurons in ferrets (Bizley & King, 2008), while studies of ventriloquism in humans (Bonath *et al.*, 2007) and monkeys (Recanzone, 1998) have implicated the auditory cortex in the shift in the perception of auditory space produced by presenting spatially-discrepant auditory-visual stimuli. Previous work in our laboratory has shown that silencing the primary auditory cortical fields in ferrets impairs approach-to-target localisation accuracy when short-duration noise bursts are used as a stimulus, without affecting head orienting behaviour (Nodal *et al.*, 2010, 2012), suggesting that the underlying neural circuits involved in guiding these responses may not be the same. This is supported by our finding that improvements in the speed and accuracy of auditory localisation

by the provision of spatially-coincident visual cues were only observed when the animals had to select which target to approach in order to receive a reward, but not when they merely turned their heads towards the target.

It is also possible that the multisensory interactions that we observed involve the visual cortex. Projections to visual areas V1 and V2 from the auditory cortex have been described in monkeys (Falchier *et al.*, 2002; Rockland & Ojima, 2003) and cats (Hall & Lomber, 2008). Intriguingly, these inputs principally target the peripheral visual field representations and, in cats, originate from the posterior auditory field, which has been shown to play a particularly significant role in sound localisation (Lomber & Malhotra, 2008). This led Hall & Lomber (2008) to suggest that these projections may serve to link cortical areas responsible for auditory and visual localisation, potentially providing an anatomical substrate for auditory cues to improve the localisation of peripheral visual targets, where spatial precision and accuracy are lowest. However, although we did indeed find that the difference between the final head bearings for visual and auditory-visual targets increased with stimulus eccentricity, this could be accounted for by the accuracy of the responses to the auditory targets. Moreover, the multisensory gain in approach-to-target response accuracy was relatively constant throughout the frontal hemifield and not restricted to peripheral targets. Further evidence against an involvement of the visual cortex in this process comes from the finding that presentation of spatiotemporally coincident visual stimuli in the blind hemifield of hemianopic humans still improves auditory localisation accuracy (Leo *et al.*, 2008), while suppression of the occipital cortex by repetitive transcranial magnetic stimulation in healthy subjects does not affect the visual enhancement of auditory localisation (Bertini *et al.*, 2010).

Elucidation of the neural circuits responsible for multisensory spatial processing requires not only the application of more refined methods for manipulating activity in specific neuronal populations (e.g. Olcese *et al.*, 2013; Wasserman *et al.*, 2015), but also measurements of the activity of those neurons during behavioural tasks. Recordings from the brain during orienting behaviour have focussed predominantly on the SC, posterior parietal cortex and, to a lesser extent, the auditory cortex and other parts of the auditory pathway and have highlighted the importance of remapping sensory representations that are used to guide movements into common reference frames (reviewed in Maier & Groh, 2009). Therefore, simultaneous recordings, while animals carry out the auditory-visual localisation task described in this study, should provide further insight into the functional significance of multisensory convergence in the auditory cortex and surrounding areas.

3.4 Appendix

3.4.1 Statistical Analyses

3.4.1.1 Final Head Bearing Errors

To investigate the effects of stimulus modality, location and duration on final head bearing errors, a linear mixed effects model was fitted to the data. A mixed effects model was used to allow for within-ferret correlation by introducing a random effect across ferrets. To optimise the model fit, error values were first converted into radians and then transformed using the sine function. The R code was as follows:

```
radhead <- (head$Final.Bearing - head$Loc)*sign(head$Loc)*pi/180
library(nlme)
mod.lme <- lme(sin(radhead) ~ Mod*Loc*I(log(Dur) - log(20)), random
= ~1|Ferret, data = head)
```

Consecutive re-levelling of the model for stimulus location and adjustment of the intercept for stimulus duration yielded the results shown in Table 3-2.

Intensity	**Final Head Bearing Error (deg)**			***P*-value (*t*-value)**		
	AV	A	V	A vs. AV	V vs. AV	A vs. V
Stimulus Location -90°						
20	-35.9	-44.3	-42.7	0	0.0001	0.3984
40	-33.7	-40.1	-42.4	0	0	0.1386

100	-30.8	-34.9	-42.0	0	0	0
200	-28.6	-31.2	-41.7	0.0039	0	0
500	-25.9	-26.5	-41.3	0.5469	0	0
1000	-23.8	-23.0	-41.0	0.4933	0	0
Stimulus Location -60°						
20	-18.1	-28.7	-30.5	0	0	0.2995
40	-17.5	-25.2	-29.5	0	0	0.0009
100	-16.6	-20.7	-28.3	0	0	0
200	-16.0	-17.4	-27.4	0.0854	0	0
500	-15.1	-13.0	-26.2	0.0188	0	0
1000	-14.5	-9.88	-25.3	0	0	0
Stimulus Location -30°						
20	-12.1	-20.2	-18.9	0	0	0.4002
40	-12.2	-17.2	-18.1	0	0	0.4968
100	-12.3	-13.4	-16.9	0.2129	0	0.0001
200	-12.4	-10.6	-16.1	0.016	0	0
500	-12.6	-6.89	-15.0	0	0.007	0
1000	-12.7	-4.10	-14.1	0	0.1827	0
Stimulus Location +30°						
20	-8.00	-7.15	-11.4	0.5307	0.0115	0.0029
40	-7.94	-7.67	-10.9	0.8057	0.0061	0.005
100	-7.86	-8.37	-10.2	0.5441	0.0036	0.0292
200	-7.81	-8.89	-9.78	0.1445	0.0089	0.2558
500	-7.72	-9.59	-9.13	0.029	0.1077	0.6035
1000	-7.67	-10.1	-8.65	0.0206	0.3664	0.1756
Stimulus Location +60°						
20	-31.1	-26.6	-36.9	0.0052	0.0003	0
40	-28.7	-25.3	-35.3	0.0079	0	0
100	-25.7	-23.7	-33.2	0.0362	0	0
200	-23.4	-22.5	-31.6	0.2566	0	0
500	-20.5	-20.8	-29.5	0.6833	0	0
1000	-18.3	-19.6	-28.0	0.2343	0	0

			Stimulus Location +90°			
20	-53.4	-52.2	-55.0	0.6275	0.4792	0.2479
40	-50.5	-49.6	-54.3	0.5792	0.0374	0.0111
100	-47.1	-46.3	-53.4	0.5163	0	0
200	-44.6	-43.9	-52.7	0.5168	0	0
500	-41.4	-40.9	-51.8	0.6279	0	0
1000	-39.2	-38.7	-51.1	0.7321	0	0

Table 3-2. Results of the linear mixed effects fitted to final head bearing error data. Modelled final head bearing errors (deg) for auditory-visual (AV), auditory (A) and visual (V) stimulus targets are shown for each stimulus location-duration combination, together with the *P*-values for the following comparisons: auditory vs. auditory-visual (A vs. AV), visual vs. auditory-visual (V vs. AV), auditory vs. visual (A vs. V).

A shapiro-Wilk test was performed and Q-Q plots (Fig. 3-11) were generated to check the fit of the linear mixed effects model. Both of these showed that the residuals were normally distributed and, therefore, demonstrated a good model fit.

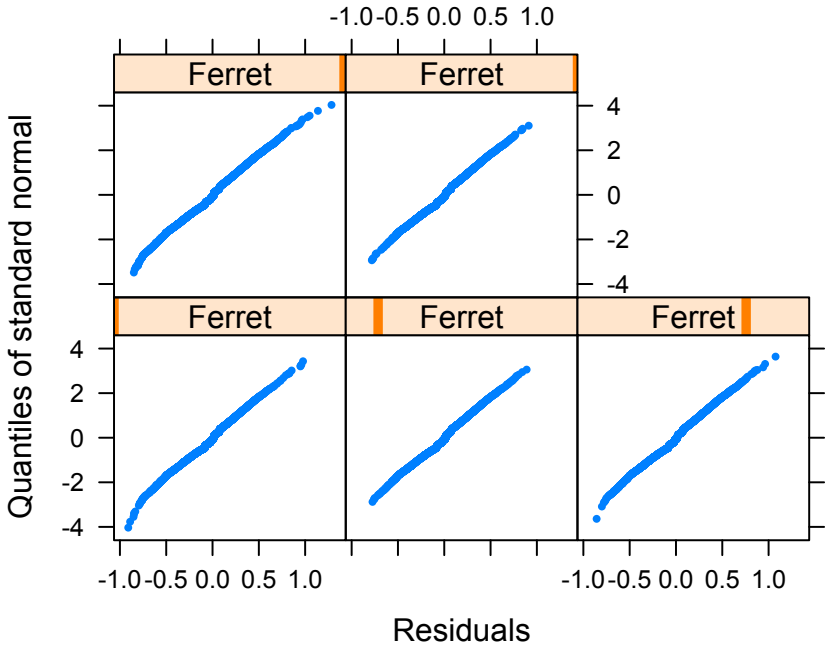


Figure 3-11. Q-Q plot of residuals from the linear mixed effects model fitted to the final

head bearing error data. Plots for individual animals are shown in separate panels. The points falling on a straight line indicate that the residuals are normally distributed.

```
shapiro.test(ranef(mod.lme)[,1])
W = 0.97375, p-value = 0.8987
qqnorm(mod.lme, ~resid(.)|Ferret, pch = 16, cex = 0.5)
```

3.4.1.2 Approach-to-Target Performance

To investigate the effect of stimulus modality, location and duration on approach-to-target response correctness, a generalised linear mixed effects model for Bernoulli distributed responses was fitted to the data. The R code used was as follows:

```
library(MASS)
mod1 <- glmmPQL(Correctness ~ Mod*Loc*I(log(Dur) - log(20)), random
= ~1|Ferret, family = binomial, data = ferret)
```

	Value	Std.Error	DF	t-value	p-value
(Intercept)	0.301	0.141	19808	2.135	0.033
ModV	-0.239	0.171	19808	-1.404	0.160
ModA	0.004	0.175	19808	0.022	0.982
Loc90	0.120	0.172	19808	0.695	0.487
Loc60	-0.675	0.170	19808	-3.967	0.000
Loc30	1.034	0.191	19808	5.400	0.000
Loc0	1.267	0.204	19808	6.217	0.000
Loc-30	0.485	0.183	19808	2.650	0.008
Loc-60	0.241	0.177	19808	1.362	0.173
I(log(Dur) - log(20))	0.761	0.076	19808	10.040	0.000

The model output shows, for example, that the log-odds corresponding to the auditory modality at -90° and 20 ms is 0.301 + 0.004. From this value it is possible to derive the probability (P_A) of a correct response to the auditory target as follows:

$$\log\left(\frac{P_A}{1-P_A}\right) = 0.004 + 0.301 = 0.305$$

$$\frac{P_A}{1-P_A} = \exp(0.305) = 1.357$$

$$P_A = 1.357(1-P_A)$$

$$P_A + 1.357P_A = 1.357$$

$$P_A = \frac{1.357}{2.357} = 0.576$$

Consecutive re-levelling of the model for stimulus location and adjustment of the intercept for stimulus duration yielded the results shown in Table 3-3.

Intensity	**Probability Correct**			***P*-value**		
	A	V	AV	A vs. AV	V vs. AV	V vs. A
Stimulus Location -90°						
20	0.576	0.515	0.604	0.982	0.160	0.151
40	0.646	0.568	0.662	0.088	0	0.012
100	0.730	0.635	0.732	0	0	0
200	0.784	0.683	0.778	0	0	0
500	0.843	0.741	0.830	0	0	0.0001
1000	0.879	0.779	0.862	0	0	0.0005
Stimulus Location -60°						
20	0.503	0.322	0.632	0.0023	0	0
40	0.578	0.388	0.718	0	0	0
100	0.672	0.482	0.81	0	0	0
200	0.735	0.555	0.863	0	0	0
500	0.806	0.647	0.914	0	0	0
1000	0.849	0.71	0.94	0	0	0

Stimulus Location -30°						
20	0.596	0.314	0.687	0.0299	0	0
40	0.64	0.384	0.757	0.0001	0	0
100	0.696	0.482	0.832	0	0	0
200	0.734	0.558	0.875	0	0	0
500	0.78	0.654	0.917	0	0	0
1000	0.81	0.719	0.94	0	0	0.0068
Stimulus Location 0°						
20	0.773	0.462	0.827	0.1129	0	0
40	0.792	0.523	0.863	0.0025	0	0
100	0.815	0.603	0.9	0	0	0
200	0.831	0.66	0.922	0	0	0
500	0.851	0.729	0.944	0	0	0
1000	0.865	0.774	0.957	0	0	0.0025
Stimulus Location 30°						
20	0.792	0.568	0.328	0	0	0
40	0.83	0.639	0.402	0	0	0
100	0.872	0.724	0.507	0	0	0
200	0.898	0.779	0.586	0	0	0
500	0.924	0.839	0.684	0	0	0
1000	0.94	0.875	0.749	0.0019	0	0
Stimulus Location 60°						
20	0.238	0.538	0.407	0.0016	0	0
40	0.314	0.609	0.525	0.0068	0	0
100	0.432	0.696	0.674	0.322	0	0
200	0.528	0.755	0.769	0.5244	0	0
500	0.651	0.819	0.861	0.0679	0	0
1000	0.733	0.858	0.909	0.0251	0	0.0001
Stimulus Location 90°						
20	0.522	0.224	0.604	0.0459	0	0
40	0.573	0.276	0.662	0.0036	0	0
100	0.639	0.355	0.732	0	0	0

200	0.686	0.421	0.778	0	0	0
500	0.743	0.512	0.83	0.0007	0	0
1000	0.78	0.581	0.862	0.0043	0	0

Table 3-3. Results of the generalised linear mixed effects model fitted to the approach-to-target response correctness data. Modelled probabilities of a correct response to auditory-visual (AV), auditory (A) and visual (V) stimulus targets are shown for each stimulus location-duration combination (left columns), together with the *P*-values (right columns) for the following comparisons: auditory vs. auditory-visual (A vs. AV), visual vs. auditory-visual (V vs. AV), auditory vs. visual (A vs. V).

To check the fit of the generalised linear mixed effects model fitted to the approach-to-target response correctness data a half normal plot was generated (Fig. 3-12). Ordered absolute values of the Pearson residuals were plotted against the expected values for a binomial distribution and, in addition, a simulated envelope was generated, which corresponds to a 95% confidence interval. The model fit is good if the plotted points fall within the boundaries of the simulated envelope.

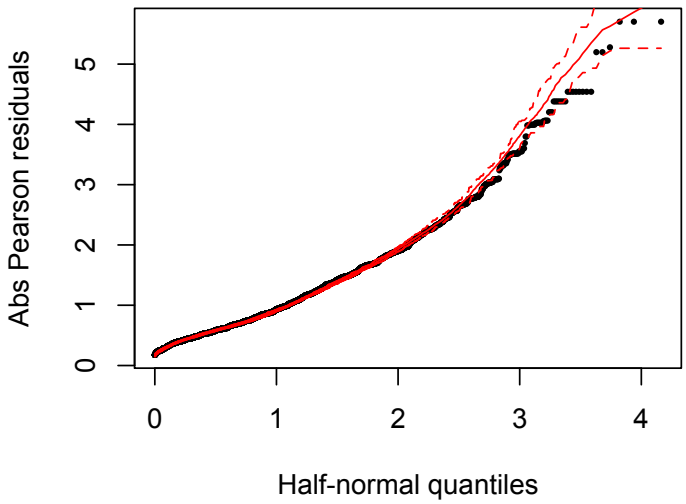


Figure 3-12. Half-normal plot of the Pearson residuals from the generalised linear model fitted to approach-to-target response correctness data. A simulated envelope (dashed lines) is also plotted together with the means of the simulated values (solid line). Most points fall within the simulated envelope, which indicates a good model fit.

3.4.1.3 Approach-to-Target Response Times

A linear mixed effects model was fitted to approach-to-target response time data to investigate the effects of stimulus modality within stimulus locations. Prior to fitting, times were first transformed using a reciprocal function. The R code used reads:

```
Library(nlme)
mod.att <- lme(1/Response_Time ~ Mod*Loc+I(log(dur)), random = ~
1|Ferret, data = att.rt)
```

Table 3-4 shows the results of the linear mixed effects model fitted to response times. The model was consecutively re-levelled to obtain results for all stimulus locations.

Location	**Response Time (s)**			***P*-value**		
	AV	A	V	A vs. AV	V vs. AV	A vs. V
-90°	1.58	1.77	1.79	0	0	0.42
-60°	1.65	1.81	1.81	0	0	0.72
-30°	1.71	1.89	1.84	0	0	0.005
0°	1.69	1.93	1.89	0	0	0.006
30°	1.67	1.91	1.86	0	0	0.003
60°	1.67	1.84	1.86	0	0	0.192
90°	1.76	1.91	1.86	0	0	0

Table 3-4. Results of the linear mixed effects fitted to approach-to-target response time data. Response times (s) to auditory-visual (AV), auditory (A) and visual (V) stimulus targets are shown (left columns) for each stimulus location, together with the *P*-values (right columns) for the following comparisons: auditory vs. auditory-visual (A vs. AV), visual vs. auditory-visual (V vs. AV), auditory vs. visual (A vs. V).

A Shapiro-Wilk test was performed and Q-Q plots (Fig. 3-13) were generated to check the fit of the linear mixed model. Both of which showed that the residuals were normally distributed and, therefore, demonstrated a good model fit.

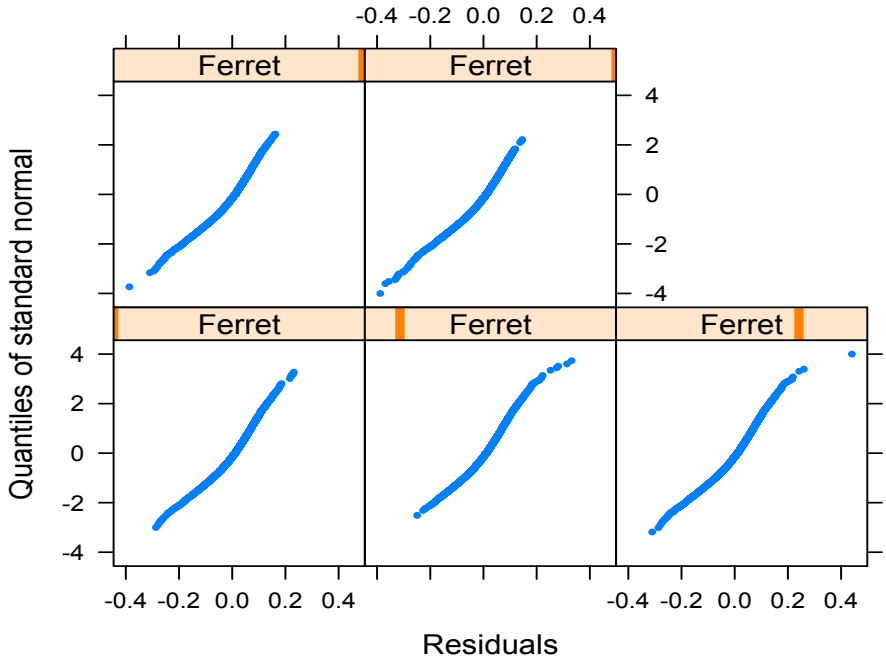


Figure 3-13. Q-Q plot of residuals from the linear mixed effects model fitted to the response time data. Plots for individual animals are shown in separate panels. The points falling on a straight line indicates that the residuals are normally distributed.

```
shapiro.test(ranef(mod.att)[,1])
W = 0.91577, p-value = 0.503
qqnorm(mod.att, ~resid(.)|Ferret, pch = 16, cex = 0.5)
```

3.4.1.4 Head Orienting Reaction Times

A linear mixed effects model was fitted to head orienting reaction time data to

investigate the effects of stimulus modality within stimulus locations. Prior to fitting, times were first transformed using a reciprocal function. The R code used reads:

```
mod.rt <- lme(1/Reaction_Time ~ Mod*Loc+I(log(Dur)), random = ~
1|Ferret, data = ho.rt)
```

Table 3-5 shows the results of the linear mixed effects model fitted to reaction times. The model was consecutively re-levelled to obtain results for all stimulus locations.

Location	**Reaction Time (ms)**			***P*-value (*t*-value)**		
(deg)	AV	A	V	A vs. AV	V vs. AV	A vs. V
-90	50.85	55.83	67.20	0.005 (-2.79)	0 (-7.30)	0 (-4.34)
-60	53.12	59.48	63.92	0.002 (-3.15)	0 (-4.43)	0.12 (-1.54)
-30	61.19	63.47	63.44	0.35 (-0.93)	0.41 (-0.82)	0.99 (0.01)
0	61.26	61.15	70.95	0.96 (0.05)	0 (-3.69)	0 (-3.60)
30	57.90	58.87	77.80	0.63 (-0.48)	0 (-6.78)	0 (-5.99)
60	54.94	57.86	84.07	0.14 (-1.44)	0 (-8.53)	0 (-7.10)
90	63.61	65.54	82.16	0.50 (-0.67)	0 (-3.94)	0 (-3.35)

Table 3-5. Results of the linear mixed effects fitted to head orienting response times. Reaction times (ms) to auditory-visual (AV), auditory (A) and visual (V) stimulus targets are shown for each stimulus location, together with the *P*-values and *t*-values (in brackets) for the following comparisons: auditory vs. auditory-visual (A vs. AV), visual vs. auditory-visual (V vs. AV), auditory vs. visual (A vs. V).

A Shapiro-Wilk test was performed and Q-Q plots (Fig. 3-14) were generated to check the fit of the linear mixed model. Both of these showed that the residuals were normally distributed and, therefore, demonstrated a good model fit.

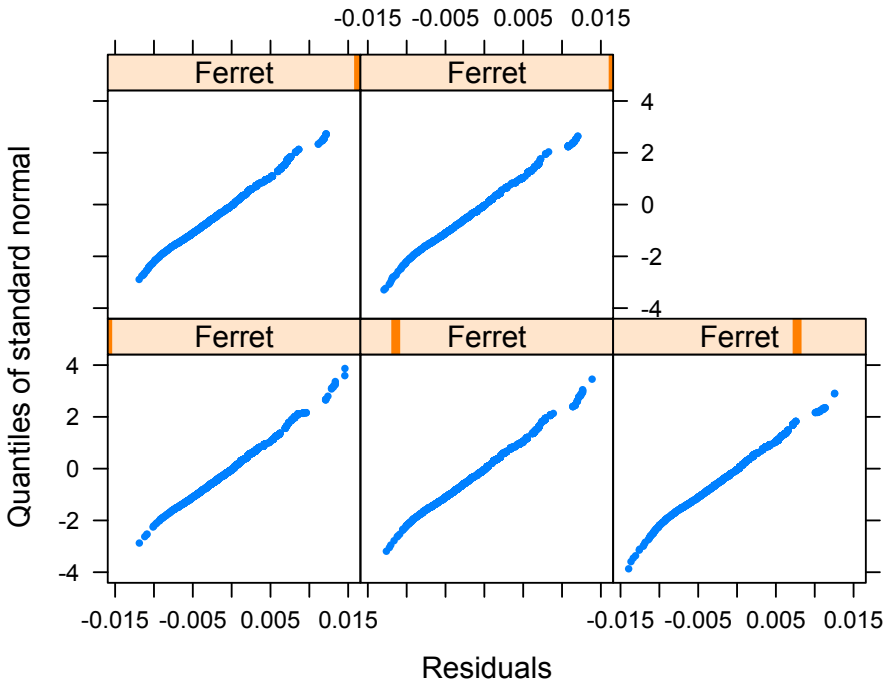


Figure 3-14. Q-Q plot of residuals from the linear mixed effects model fitted to the reaction time data. Plots for individual animals are shown in separate panels. The points falling on a straight line indicates that the residuals are normally distributed.

```
shapiro.test(ranef(mod_rt)[,1])
W = 0.97859, p-value = 0.927
qqnorm(mod.rt, ~resid(.)|Ferret, pch = 16, cex = 0.5)
```

4 Neural Activity in the Auditory Cortex during Multisensory Localisation

4.1 Introduction

It is increasingly clear that focusing exclusively on the responses of neurons to the acoustic properties of sound is insufficient if we wish to understand how activity in the auditory system, and the cortex in particular, underpins perception and behaviour in multisensory environments. The use of more naturalistic stimuli to study auditory processing has shown that multisensory interactions are a ubiquitous feature of the auditory system (reviewed by Schroeder & Foxe, 2005; Ghazanfar & Schroeder, 2006). For example, at the level of the auditory cortex, studies have shown that aside from the modulation of activity by visual stimulation, activity can be driven by visual stimulation alone (monkeys: Schroeder & Foxe, 2002; Brosch *et al.*, 2005;

Ghazanfar *et al.*, 2005; Bizley & King, 2008; Kayser *et al.*, 2008; ferrets: Meredith & Allman, 2015). Although multisensory interactions have been described even at the level of the core auditory cortex (Brosch *et al.*, 2005; Ghazanfar *et al.*, 2005; Bizley & King, 2008), they have been studied predominantly in anaesthetised or passively listening animals and, therefore, their effect on the primary role of the auditory cortex in hearing and, ultimately, their behavioural relevance is relatively unknown.

A growing body of experimental evidence suggests that the auditory cortex may provide a substrate for visual influences on spatial hearing. First, an intact auditory cortex is required for normal spatial hearing (Malhotra & Lomber, 2007; Nodal *et al.*, 2010). Second, the auditory cortex receives inputs from a region of extrastriate visual cortex that has been implicated in spatial processing (Boehnke & Phillips, 2005; Bizley *et al.*, 2007). Third, the presence of spatially congruent visual stimuli has been shown to produce an overall gain in the spatial information available in the responses of auditory cortical neurons in the anaesthetised ferret (Bizley & King, 2008). However, whereas the superior colliculus (SC) contains topographically-aligned maps of sensory space (King & Hutchings, 1987), it is less obvious how visual inputs are coordinated with auditory representations that, at least in early levels of cortical processing, are tonotopically organised.

Elucidating the neural circuits responsible for multisensory spatial processing is particularly difficult, since this field of research has yet to benefit to any great degree from the experimental approaches that are now available for interrogating the functions of specific neural circuits (Olcese *et al.*, 2013; Wasserman *et al.*, 2015). For example, neurophysiological recordings in awake-behaving animals enable robust inferences to be made about the link between the activity of single or

populations of neurons and the functional significance of the processes being investigated; however they have been applied predominantly to the study of individual sensory systems, including vision (Britten *et al.*, 1992, 1996; Leopold & Logothetis, 1996; Yang & Maunsell, 2004), audition (Recanzone, Guard, & Phan, 2000; Recanzone, Guard, Phan, *et al.*, 2000; Niwa *et al.*, 2012a; Bizley *et al.*, 2013) and touch (Steinmetz *et al.*, 2000; Romo *et al.*, 2002). In comparison, relatively few studies have attempted to correlate multisensory interactions at the neural and behavioural levels by recording neural activity during the performance of a multisensory task, with those that have, focusing predominantly on the SC (Frens & Van Opstal, 1998; Bell & Corneil, 2003; Opstal & Munoz, 2004), visual cortex (Wang *et al.*, 2008), higher auditory cortex (Chandrasekaran *et al.*, 2013) and prefrontal cortex (Hwang & Romanski, 2015), rather than the early auditory cortex. Therefore, in the present study, neural activity from early processing areas of the ferret auditory cortex has been recorded during the performance of the localisation task, to investigate the extent and functional significance of multisensory convergence in this region. The results presented in this chapter also provide insight into how sound source location is encoded by populations of cortical neurons.

4.2 Results

Five ferrets were trained in a 2-alternative forced choice (2-AFC) task to localise auditory, visual and auditory-visual stimuli presented pseudo-randomly from either the left (-30°) or the right (+30°) hemifield. In order to explore the effect of the visual stimulus on auditory processing and based on the target region for the neural recordings (i.e. the auditory cortex) sound level was roved pseudo-randomly across trials in 7 dB steps over either 69 – 90 dB SPL (F1305) or 20 – 55 dB SPL (F1324, F1404, F1440 and F1441), whilst the luminosity of the visual stimulus was kept constant across trials at 17 cd. Also, in order to ensure that animals did not reach near perfect performance, the duration of stimuli was fixed at 100 ms. Neural activity was recorded simultaneously during behavioural task performance from multi-electrode arrays implanted bilaterally over the auditory cortex. The results presented in this chapter are based on the analysis of the neural responses recorded over 26,346 trials (see Table 2.1 for a breakdown of the number of trials per animal).

4.2.1 Task Performance During the Neural Recordings

Psychometric curves were fitted to behavioural data obtained during the neural recording sessions to relate the percentage of trials on which animals made a right-sided (+30°) approach-to-target response to the intensity and location of the stimulus. The curves obtained from all ferrets show that for both auditory and auditory-visual stimuli, the percentage of correct right- or left-sided responses increased with

increasing sound intensity (Fig. 4-1 and Appendix 4.4.1). For two animals (F1324 and F1404), the fitted psychometric curves do not correspond at the interpolated midline (0 dB SPL) to the 50% choice probability point (Table 4-1). These small shifts in the psychometric curves indicate a response bias, i.e. when presumably unsure about the correct response, animals do not make right or left responses with exactly the same probability, but instead show a preference for a particular reward spout. In both cases, the values at 50% were negative indicating a right sided bias, which is consistent with previous reports in this species (Nodal *et al.*, 2008; Walker *et al.*, 2009). The ability of ferrets to localise unisensory and multisensory stimuli was compared by the calculation of two parameters: the slopes of the fitted psychometric curves and the intensity values corresponding to threshold performance (defined as 75%). A comparison of slope magnitudes shows that for all animals, values were higher for auditory-visual stimuli (slope ± SEM: 1.5 ± 0.2) than for auditory stimuli (slope ± SEM: 1.2 ± 0.2) (for individual values see Table 4-1, right columns). Accordingly, threshold values were also smaller for auditory-visual stimuli (16.7 ± 2.6 dB SPL) than for auditory stimuli (21.1 ± 2.9 dB SPL) (Table 4-1). Together, the higher slope magnitudes and lower detection thresholds demonstrate that localisation responses were more accurate to auditory-visual targets than to auditory targets, which is consistent with the results observed for the localisation task described in Chapter 3.

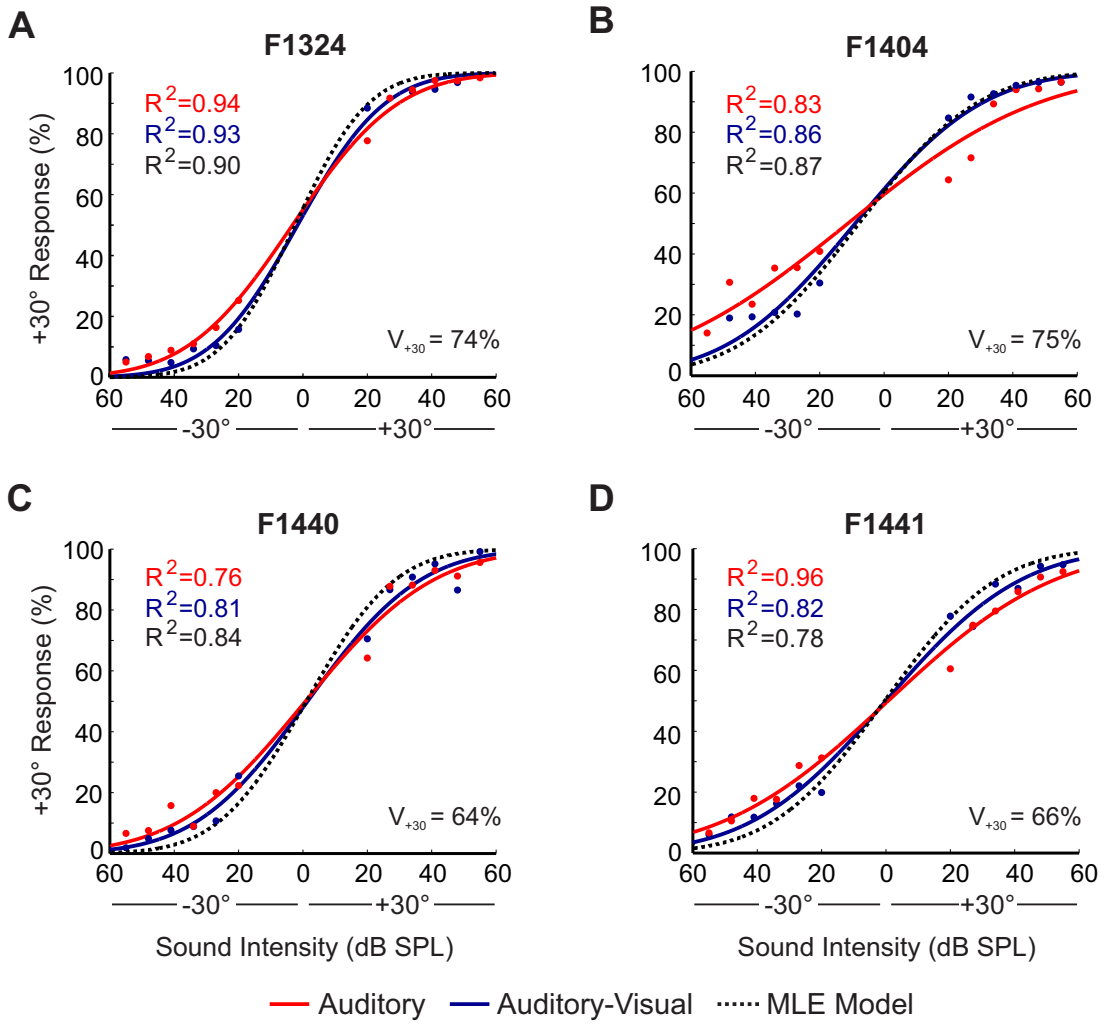


Figure 4-1. Psychometric curves showing approach-to-target performance during the neural recordings. (A-D) The percentage of right-sided (+30°) approach-to-target responses is shown as a function of the stimulus intensity and location for the auditory stimulus either presented alone (Auditory, red dots) or with a spatially coincident visual stimulus (Auditory-Visual, blue dots). Data are fitted with a cumulative Gaussian distribution function (Auditory, red lines; Auditory-Visual, blue lines). On each panel R^2 values of the fits are shown, together with the percentage of correct responses to the visual stimulus at +30° (V_{+30}). Modelled auditory-visual psychometric curves derived using maximum likelihood estimation (MLE, black dotted lines) are also shown, together with the corresponding R^2 values for the model fit to the measured auditory-visual percentage values.

	Intensity Value 50%			Intensity Value 75%			Slope Magnitude		
	A	AV	MLE	A	AV	MLE	A	AV	MLE
F1324	-3.3	-1.7	-2.5	14.1	12.7	9.6	1.6	1.9	2.2
F1404	-11.3	-9.0	-7.8	20.3	12.1	11.8	0.9	1.3	1.4
F1440	0.8	1.4	1.1	21.8	20.0	15.7	1.3	1.4	1.8
F1441	0.8	-0.4	-0.4	28.2	22.3	18.1	1.0	1.2	1.5

Table 4-1. Psychometric curve parameters. Psychometric curves were fitted to relate the percentage of trials on which animals made a right-sided (+30°) approach-to-target response to stimulus intensity and location. For each animal (rows), intensity values corresponding to the 50% and 75% probability points are shown along with the slopes of curves fitted to auditory (A), auditory-visual (AV) and modelled (MLE) data.

To explore whether ferrets integrate auditory-visual information in a statistically optimal manner, a Maximum Likelihood Estimation (MLE) method was employed to model the auditory-visual psychometric curves (Fig. 4-1 A-D, black dashed lines) (see Section 2.6.4 for method). This method predicts the behavioural performance expected for the auditory-visual stimulus condition if sensory inputs are weighted by their relative reliabilities and integrated in a linear fashion. In all cases, slope magnitudes were higher and threshold values were lower for the modelled curves (slope ± SEM: 1.7 ± 0.2; intensity ± SEM: 13.8 ± 1.9) than for the empirical curves (slope ± SEM: 1.5 ± 0.2; intensity ± SEM: 16.7 ± 2.6) (for individual values see Table 4-1). This illustrates that the model slightly overestimates the multisensory enhancement effect on approach-to-target performance. However, in all cases, comparison of R^2 values for the modelled and empirical psychometric curve fits to the measured audio-visual percentage values shows that the absolute difference is less than 5% (F1324: 3%; F1404: 0.7%; F1440: 4.26%; F1441: 3.7%). Thus, task

performance closely approximates that predicted by the model and suggests that neural integration in the ferret occurs in a fashion compatible with the MLE principle.

A generalised linear model for Bernoulli-distributed responses was fitted to the approach-to-target data to investigate the significance of the multisensory enhancement effect on task performance (see Fig. 4-2A-B and appendix 4.4.2.1 for significance values). For most stimulus location-intensity combinations, the model shows that animals were significantly ($P < 0.05$) more likely to make a correct approach-to-target response to auditory-visual stimuli than to auditory stimuli, with the exception of +30° at the highest sound intensities tested (48-55 dB SPL). Therefore, despite reducing the parametric dimensionality of the stimuli from that used in Chapter 3, our behavioural results provide robust evidence for neural integration. Thus, validating the use of the modified version of the task, in combination with the neural recordings, to investigate the relationship between multisensory activity in the auditory cortex and the crossmodal enhancement of localisation behaviour.

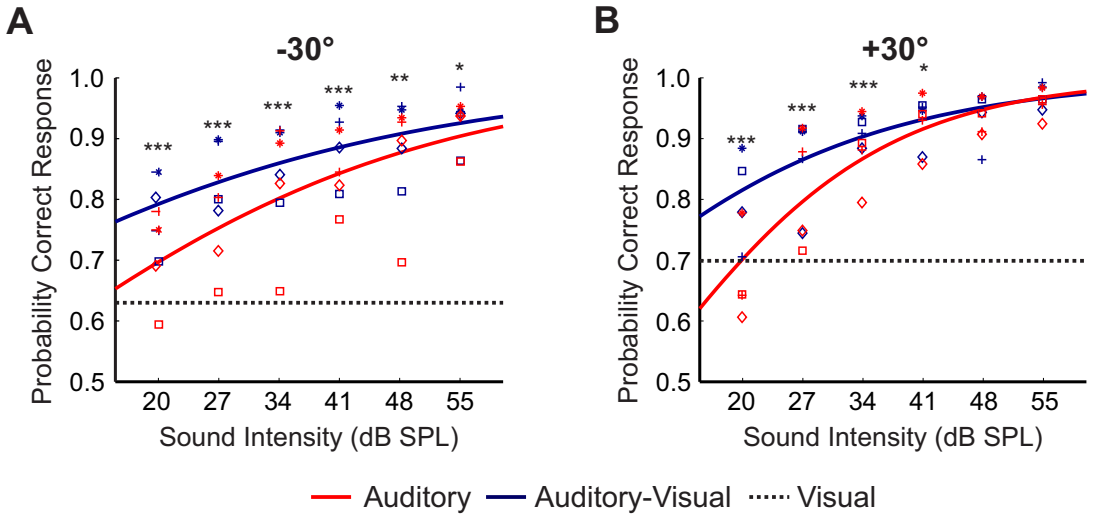


Figure 4-2. Results of the generalised linear mixed effects model fitted to the approach-to-target behavioural data. Modelled probability values of a correct approach-to-target response are shown as a function of sound intensity for auditory (red) and auditory-visual (blue) stimulus targets. Results for each target location are shown in the separate panels. Data from individual animals are represented by the different symbols. Significant differences between stimulus conditions are indicated by asterisks (*P < 0.05; ** P < 0.01 and *** P < 0.001). For comparison, the probability of a correct response to the fixed brightness visual stimulus is shown by the black dotted line.

4.2.2 Types of Neural Responses Evoked During the Localisation Task

For each unit, responsiveness to auditory and/or visual stimuli was established by fitting a generalised linear model for a Poisson distribution to the data obtained from the localisation task (see Section 2.6.5 for full details). In total, 627 single units were identified as responsive to auditory and/or visual stimulation using this method (Table 4-2A) and confirmed, by histological analysis, to be located within the ectosylvian gyrus. The aim of this study was to explore multisensory integration in early auditory processing cortices. Therefore, electrode arrays were implanted over the MEG and the dorsal PEG (Fig. A-1). This region was chosen because it typically contains neurons that both exhibit characteristics associated with early cortical

processing areas, including short auditory response latencies and frequency sensitivity (Kelly & Judge, 1994; Wallace *et al.*, 1997; Bizley *et al.*, 2005), and display some sensitivity to visual stimulation (Bizley *et al.*, 2007). For a full description of the recording site locations for each animal see Appendix A.

The majority (584/627) of units responded to auditory stimulation. Of these 492/584 exhibited excitatory and 92/584 inhibitory responses. In addition, a smaller proportion (116/627) of units responded to visual stimulation, of which 65/116 showed excitatory and 51/116 inhibitory responses (see Table 4-2A). Auditory minimum response latencies were calculated for all units that produced an excitatory response to sound presented alone (n=492) and visual minimum response latencies were calculated for all units that produced an excitatory response to light presented alone (n=65). Overall, visual latencies (mean ± SEM: 57.0 ± 5.0 ms) were almost twice as long as compared to auditory latencies (mean ± SEM: 27.4 ± 1.1 ms).

All units that exhibited an excitatory response to auditory and/or visual stimulation (509/627) were classified as auditory, visual or bimodal (Table 4-2B). The latter type included units that were driven either by both modalities when presented independently (bimodal supra-threshold) or by only one modality but whose responses were significantly modulated (either enhanced or suppressed) by the simultaneous presentation of the second, otherwise ineffective, modality (bimodal sub-threshold). Therefore, for all units that were classified by the generalised linear model as responsive to only one modality, a further comparison was performed within the model to identify whether its response was significantly modulated by the simultaneous presence of the second modality, which when presented independently failed to elicit a response.

According to this classification, the majority of units from the region sampled were classified as unimodal auditory (84%, 428/509). Nevertheless, at the single cell level, the activity of some neurons was significantly modulated and sometimes even driven by visual inputs. In total, 16% (81/509) of units were found to be influenced by visual stimulation (i.e. classified as either unimodal visual (23), bimodal sub-threshold (22) or bimodal supra-threshold (36)), as summarised in Table 4-2B and across animals in Figure 4-3. The proportion of units sensitive to visual stimulation is consistent with those values previously reported for the anaesthetised ferret auditory cortex (Bizley *et al.*, 2007; Bizley & King, 2009; Meredith & Allman, 2015). Across animals, the proportion of units that were found to be sensitive to visual stimulation varied from 9% (F1305) to 36% (F1441); this difference broadly mirrors the variation in cortical field location of implants (see Appendix A).

A. Stimulus Responsiveness

	A_{Unres}	A_{Excit}	A_{Inhib}	Total
V_{Unres}	-	445	66	511
V_{Excit}	28	36	1	65
V_{Inhib}	15	11	25	51
Total	43	492	92	627

B. Unit Classification

	Unimodal		Bimodal		Total
	Auditory	Visual	Supra	Sub	
N	428	23	36	22	509
%	84	5	7	4	100

Table 4-2. Summary of unit responsiveness and classification. (A) Number of units responsive to auditory and/or visual stimulation. Abbreviations: A_{Unres} auditory unresponsive; A_{Excit} auditory excitatory; A_{Inhib} auditory inhibitory; V_{Unres} visual unresponsive; V_{Excit} visual excitatory; V_{Inhib} visual inhibitory. (B) Number (N) and percentage (%) of units classified as unimodal auditory, unimodal visual, bimodal supra-threshold (Supra) or sub-threshold (Sub). Note that only units that produced an excitatory response were classified (shaded in grey). Units are pooled across animals.

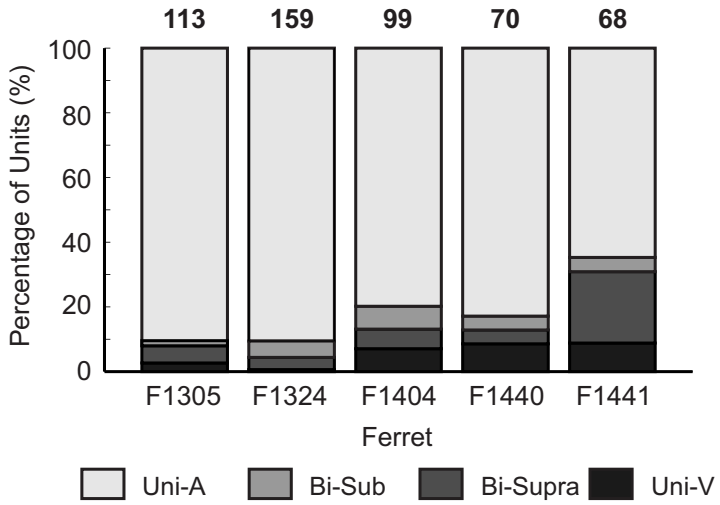


Figure 4-3. Breakdown of unit types across animals. The percentage of units classified as each type is shown for each animal, together with the total number of units displayed above each bar. Unit types included: unimodal auditory (Uni-A), unimodal visual (Uni-V), bimodal suprathreshold (Bi-Supra) and bimodal subthreshold (Bi-Sub).

To highlight the range of responses recorded from the auditory cortex during the task, the responses of a number of example units are illustrated in Figures 4-4, 4-5 and 4-6. The raster plots and histograms in Fig. 4-4A-B show examples of units that responded only to auditory stimulation (i.e. visual stimulation either in isolation or in combination with auditory stimulation failed to change their firing rates). According to the temporal course of their responses, the first unit shows an example of an onset response (Fig. 4-4A, i.e. the maximum response occurred in the 50 ms after stimulus onset). In primary auditory cortex neurons, auditory stimuli typically evoke onset responses whereby a brief depolarization is followed by a period of hyper-polarisation due to an increase in inhibitory synaptic conductance (Tan *et al.*, 2004). The second unit shows an example of an offset response (Fig. 4-4B, i.e. the maximum response occurred in the 50 ms after stimulus offset). In contrast, some

units were driven by both auditory and visual stimulation, as illustrated by the two units shown in Figure 4-5. The first unit shows an example of an excitatory response (Fig. 4-5A), whereas the second shows an example of an inhibitory response (Fig. 4-5B). Finally, Figure 4-6 shows two examples of bimodal subthreshold units, which were driven only by auditory stimulation but their responses were modulated, either enhanced (Fig. 4-6A) or depressed (Fig. 4-6B), by simultaneous visual stimulation.

The classification of units as unimodal or bimodal is useful from the perspective that it enables comparisons to be drawn between studies regarding the extent of cross-modal influences. However, recent studies have shown that neurons in the auditory cortex (Meredith & Allman, 2015) and extrastriate visual cortex (Allman *et al.*, 2009) that are classified as unimodal, at the single cell level, display subtle crossmodal effects that are only detectable at the population level. Therefore, in order to investigate comprehensively the effects of visual stimulation on auditory cortical activity, the subsequent analysis of the effect of multisensory interactions on the rate, timing and regularity of spike discharge patterns was performed not only for units classified as bimodal but also for those classified as unimodal auditory. In addition, although the majority of units exhibited an onset response (mean auditory response latency: 27 ms), some units exhibited an offset response with a similar latency to the onset response from the stimulus offset. This precise temporal pattern of evoked activity, along with temporal disparity between auditory and visual inputs determined the time window (150 ms) that was used for the subsequent analysis (see Section 2.6.5).

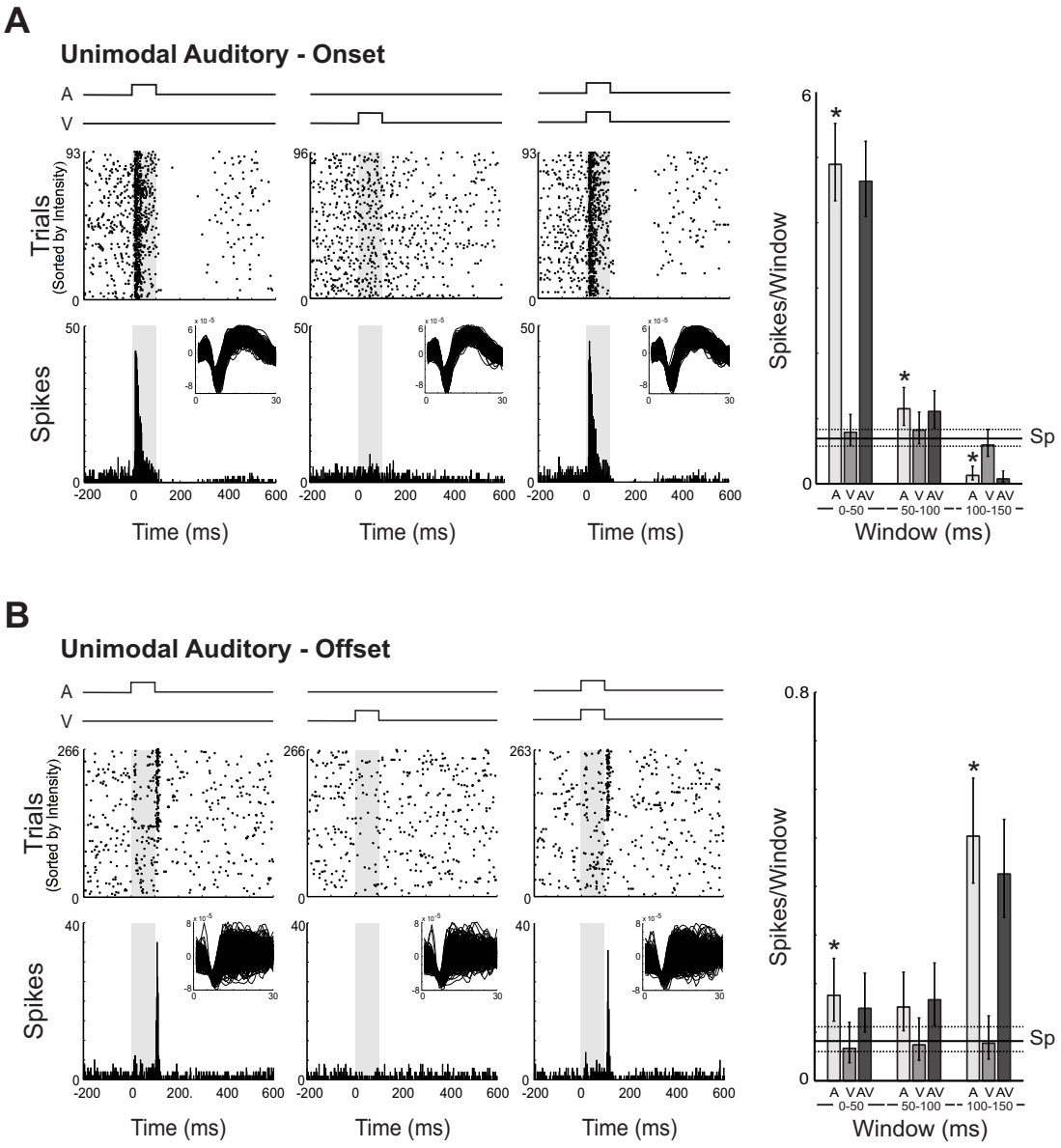


Figure 4-4. Example unimodal auditory units exhibiting an onset (A) or offset (B) response. Single-unit recordings from neurons in the auditory cortex in response to auditory (A), visual (V) and audio-visual (AV) stimuli of 100 ms duration. Each panel represents data from a different single unit and shows stimulus traces (top row), raster plots (middle row, one dot = one spike, one row = one trial), peri-stimulus time histograms (bottom row, 2 ms time bins) and spike waveforms (bottom row inset). A bar chart (far right) is also shown, summarising the mean (± 99% CI) number of spikes for each stimulus type during three separate time windows (0 – 50, 50 – 100 and 100 – 150 ms from the stimulus onset). The horizontal line represents the spontaneous (Sp) activity level.

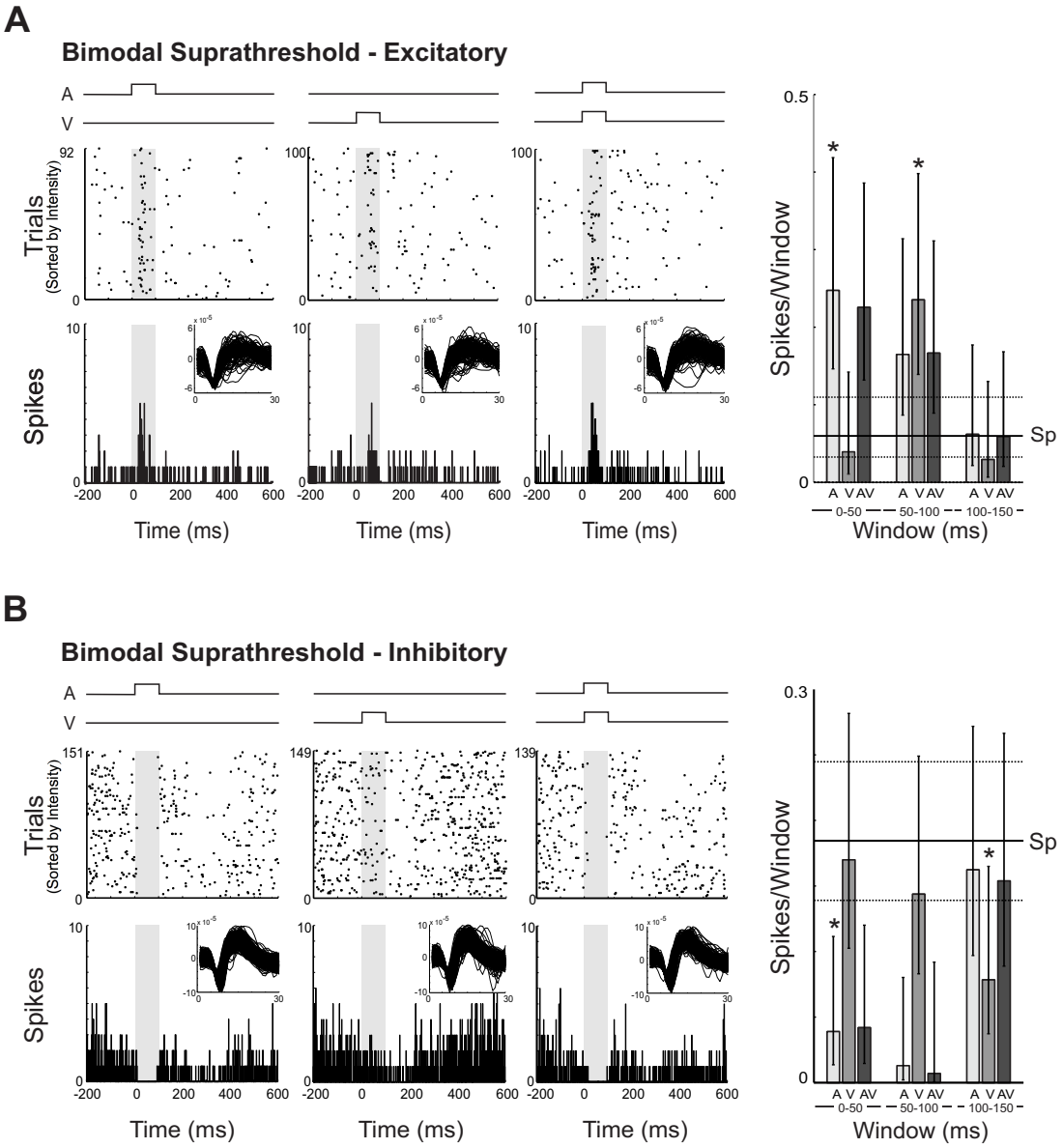


Figure 4-5. Example bimodal supra-threshold units exhibiting an excitatory (A) or inhibitory response (B). For a detailed description of panels see Figure 4-4 legend.

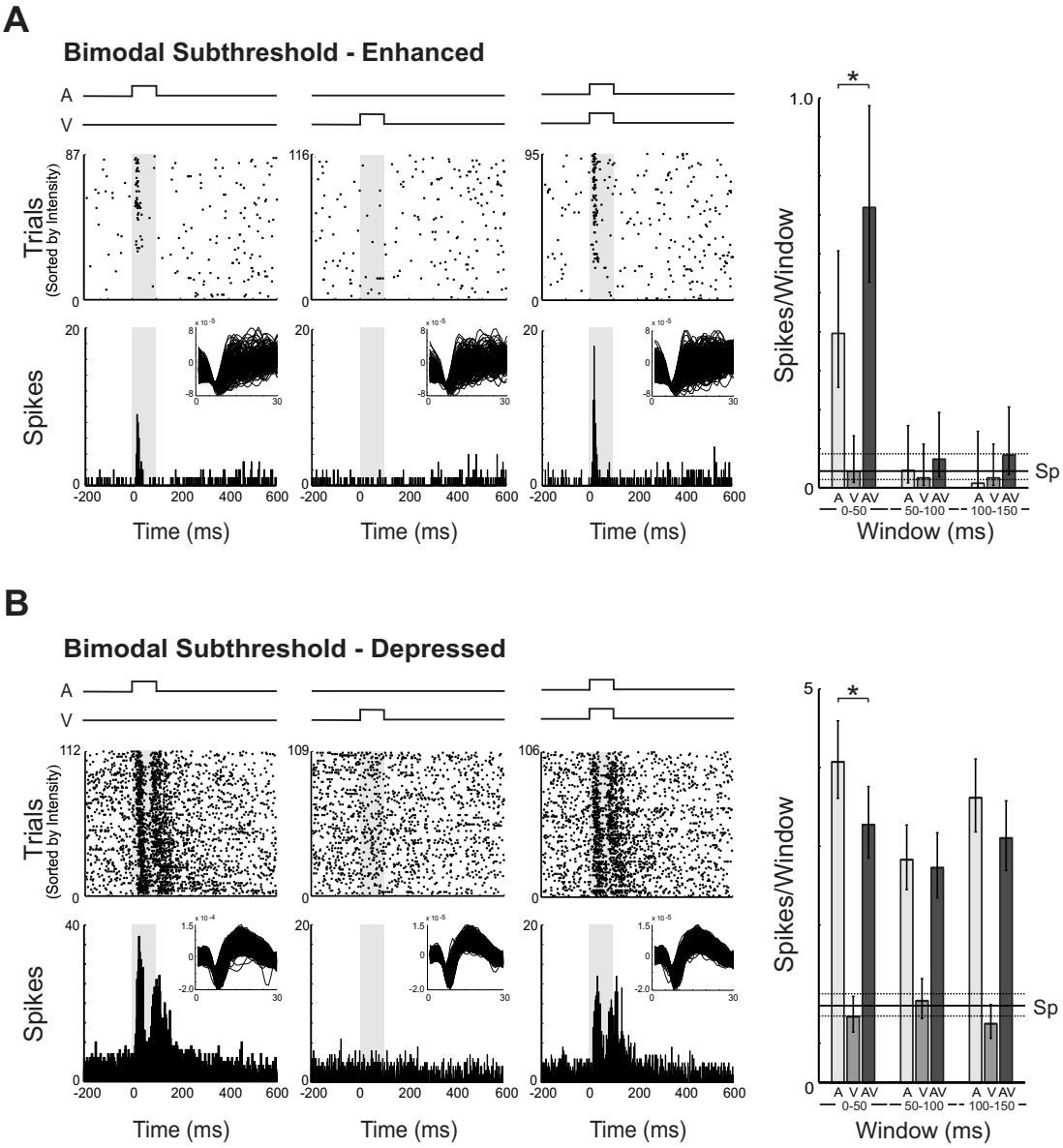


Figure 4-6. Example bimodal sub-threshold units showing enhanced (A) or suppressed (B) modulation. For a detailed description of panels refer to Figure 4.4 legend.

4.2.3 Basic Response Properties of Neurons

In order to sample as many visually responsive units in the auditory cortex as possible the intensity of the visual stimulus was kept constant (white light LED of 17 cd), at a level which has been shown to drive spiking activity in the anaeasthetised ferret auditory cortex (Bizley *et al.*, 2007), whereas sound level was varied. On average, auditory and bimodal units had similar thresholds (mean ± SEM: 31.7 ± 0.7 and 32.8 ± 2.5 dB SPL, respectively) for the auditory stimulus. In comparison, thresholds for the auditory-visual stimulus were marginally higher for both auditory and bimodal units (mean ± SEM: 31.8 ± 2.5 and 36.5 ± 2.5 dB SPL, respectively).

The effects of stimulus location, modality and intensity on neural activity were explored by generating scatter plots and by fitting a generalised linear model for a negative binomial distribution to the spike count data (i.e. the number of spikes recorded in a 150 ms window after stimulus onset). On average, contralateral auditory stimuli evoked stronger responses than ipsilateral auditory stimuli in both unimodal auditory (Contra vs. Ipsi 1.06 ± 0.06 vs. 0.79 ± 0.05 spikes/trial) and bimodal units (Conta vs. Ipsi 1.32 ± 0.29 vs. 1.03 ± 0.25 spikes/trial). This contralateral preference is shown in Figures 4-7A and B, which show that most units lie above the unity line on scatter plots comparing contralateral to ipsilateral evoked activity, and is consistent with previous reports in a variety of species, including ferrets (Mrsic-Flogel *et al.*, 2001, 2003, 2005; Campbell *et al.*, 2010), cats (Middlebrooks & Pettigrew, 1981; Imig *et al.*, 1990; Rajan *et al.*, 1990; Brugge *et al.*, 1996), rats (Yao *et al.*, 2013) and monkeys (Miller & Recanzone, 2009). For both unit types, this effect was significant ($P < 0.001$) across the majority of sound

intensities tested (see Fig. 4-7C-D and appendix 4.4.2.2 for significance values).

Previous studies have suggested that one function of visual input to the auditory cortex might be to sharpen the relatively imprecise spatial tuning of neurons (Bizley *et al.*, 2007; Bizley & King, 2008, 2009). Although in the present study we were unable to examine the effect of concomittant visual stimulation on the exact size of the spatial receptive fields, since only two stimulus locations were used, we were able to examine its effect on the preference of units to contralateral space. As for the auditory stimulus, contralateral auditory-visual stimuli evoked stronger responses than ipsilateral auditory-viusal stimuli in both unimodal auditory (Contra vs. Ipsi 1.05 ± 0.06 vs. 0.76 ± 0.06 spikes/trial) and bimodal units (Contra vs. Ipsi 1.19 ± 0.24 vs. 0.94 ± 0.23 spikes/trial) (Fig. 4-7E-F). For both unit types, this effect was significant ($P < 0.001$) across all of the sound intensities tested (Fig. 4-7G-H and appendix 4.4.2.2 for significance values). A comparison across auditory and auditory-visual stimulus types of the slopes and y-intercepts of the regression lines fitted to the scatter plots illustrates that the addition of the visual stimulus had no noticable effect on the contralateral preference of units. This was the case regardless of whether units were classified as unimodal auditory or bimodal (Table 4-3). Although we cannot exclude the possibility that concomitant visual stimulation might have refined the spatial sensitivity of the neurons as described in anaesthetised animals (Bizley *et al.*, 2007), the lack of change in the contralateral preference suggests that other regions may contribute to the localisation performance gain.

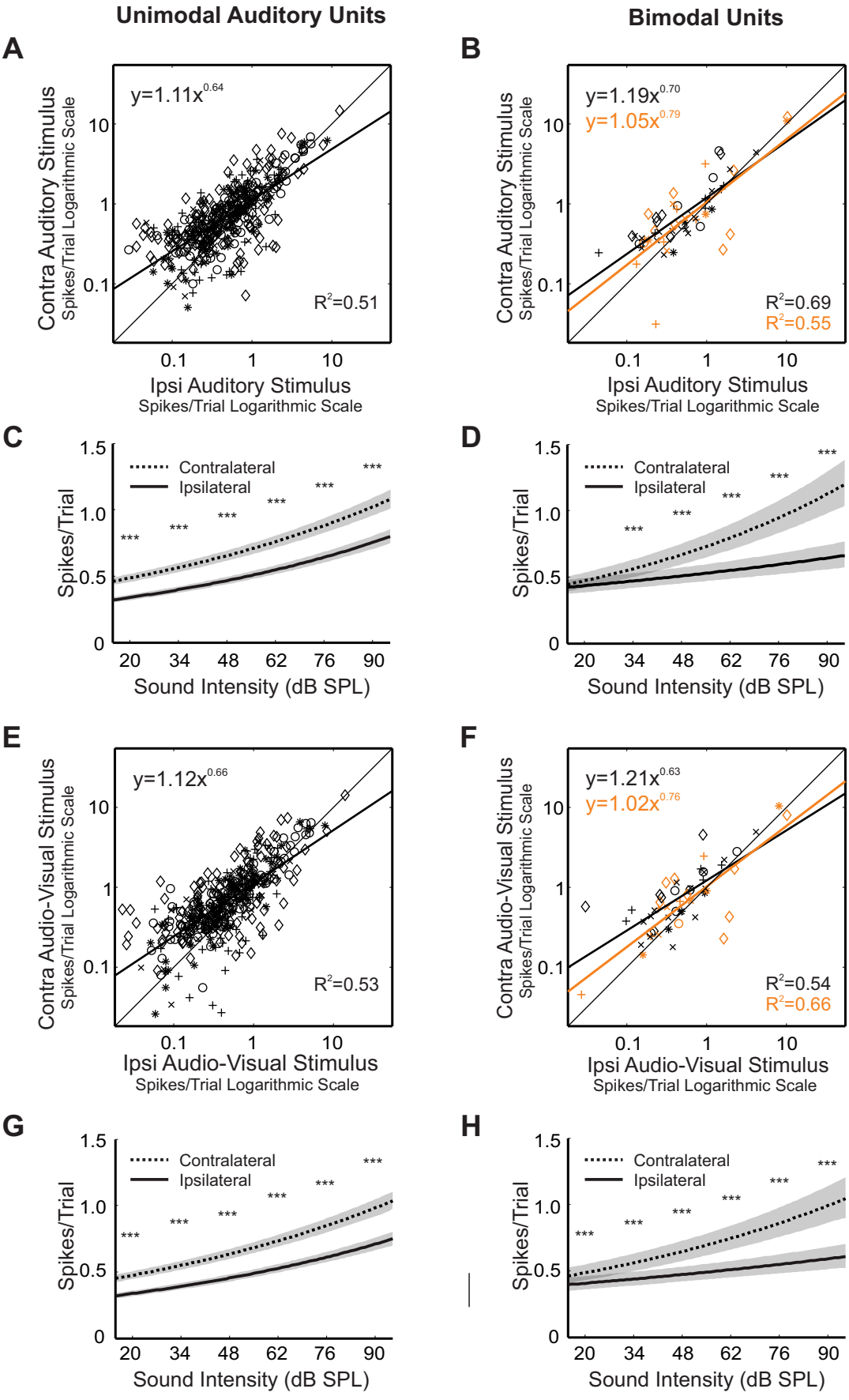

Unimodal Auditory Units
Bimodal Units
A
B
C
D
E
F
G
H
y=1.11x^0.64
R²=0.51
y=1.19x^0.70
y=1.05x^0.79
R²=0.69
R²=0.55
Contra Auditory Stimulus
Spikes/Trial Logarithmic Scale
Ipsi Auditory Stimulus
Spikes/Trial Logarithmic Scale
Contralateral
Ipsilateral
Spikes/Trial
Sound Intensity (dB SPL)
y=1.12x^0.66
R²=0.53
y=1.21x^0.63
y=1.02x^0.76
R²=0.54
R²=0.66
Contra Audio-Visual Stimulus
Spikes/Trial Logarithmic Scale
Ipsi Audio-Visual Stimulus
Spikes/Trial Logarithmic Scale

Figure 4-7. Effect of stimulus location on spike counts. (A-B) Scatter plots comparing the number of spikes evoked by contralateral and ipsilateral auditory stimuli in (A) unimodal auditory and (B) bimodal units (suprathreshold units, black; subthreshold units, orange). Points above the unity line indicate that units responded more strongly to contralateral than to ipsilateral stimuli. A regression line was fitted to each scatter plot, using the linear least squares method, and the corresponding slope and y-intercept are shown along with the R^2 value in each panel (see Table 4-4 for a complete set of values). Data from individual animals are represented by different symbols. (C-D) Corresponding modelled spike count values are shown as a function of sound intensity for contralateral auditory (dotted line) and ipsilateral auditory (black line) stimuli below each scatter plot. Significant differences between stimulus conditions are indicated by asterisks and *P*-value by their number: * $P < 0.05$; ** $P < 0.01$ and *** $P < 0.001$). (E-F) Scatter plots comparing the number of spikes evoked by contralateral and ipsilateral auditory-visual in (E) unimodal auditory and (F) bimodal units (suprathreshold units, black; subthreshold units, orange). (G-H) Corresponding modelled spike count values are shown as a function of sound intensity for contralateral auditory-visual (dotted line) and ipsilateral (black line) auditory-visual stimuli.

	Unimodal Auditory		Bimodal Sub		Bimodal Supra	
	A	AV	A	AV	A	AV
Slope	0.64	0.66	0.78	0.76	0.70	0.63
Y-intercept	1.11	1.12	1.05	1.03	1.19	1.22

Table 4-3. Parameters for the regression lines fitted to the spike count scatter plots. The slopes and Y-intercepts are shown for regression lines fitted to scatter plots comparing the number of spikes evoked per trial by contralateral and ipsilateral stimuli. Values are shown separately for auditory (A) and auditory-visual (AV) stimuli and for unimodal auditory (left columns), bimodal subthreshold (Bimodal Sub, middle columns) and bimodal suprathreshold (Bimodal Supra, right columns) units.

As might be expected for unimodal auditory units, the number of spikes evoked per trial by the auditory stimulus was substantially higher than the number recorded following presentation of the visual stimulus (A vs. V 1.06 ± 0.06 vs. 0.44 ± 0.03 spikes/trial) (Fig. 4-8A, C), which was comparable to the spontaneous activity (V vs. Spon 0.44 ± 0.03 vs. 0.46 ± 0.03 spikes/trial). For bimodal units, the responses evoked by auditory stimuli were stronger than those evoked by visual stimuli (Fig. 4-

8B, D). However, in keeping with their classification, this difference was less pronounced for suprathreshold units (A vs. V 1.06 ± 0.19 vs. 0.78 ± 0.13 spikes/trial) than for subthreshold units (A vs. V 1.88 ± 0.76 vs 0.87 ± 0.37 spikes/trial).

For units driven by visual stimulation (i.e. classified as visual or bimodal suprathreshold) visual-evoked activity was, as expected, higher than the spontaneous activity (V vs. Sp, visual units: 0.67 ± 0.08 vs. 0.46 ± 0.07 spikes/trial). However, this activity was substantially lower, in terms of evoked activity expressed as a percentage increase of spontaneous activity, than auditory-evoked activity: 42% vs. 113%, respectively. In addition, we found that for units driven by visual stimulation, contra- and ipsilateral visual stimuli evoked similar responses (Contra vs. Ipsi, visual units: 0.67 ± 0.09 vs. 0.63 ± 0.08 spikes/trial). This was in contrast to units driven by auditory stimulation, which showed a clear preference for contralateral stimuli.

Interestingly, when the unimodal auditory units were considered together, the presence of a visual stimulus subtly suppressed the mean auditory response of the population (A vs. AV 1.06 ± 0.06 vs 1.04 ± 0.06 spikes/trial) (Fig. 4-8E). Likewise, a similar suppressive effect was found for bimodal suprathreshold units (A vs. AV 1.06 ± 0.19 vs 1.04 ± 0.18 spikes/trial) and subthreshold units (A vs. AV 1.82 ± 0.73 vs 1.51 ± 0.58 spikes/trial) (Fig. 4-8F). It is perhaps not suprising, given that visually responsive units were not tuned to a particular side of space, that this suppressive effect was evident for both contra- and ipsilateral stimuli (data only shown for contralateral stimuli). In total, visual co-stimulation suppressed auditory-evoked activity by 3.2% across the entire population of units driven by auditory stimulation. This suppressive effect, although subtle, was significant for both unimodal auditory and bimodal units (see Fig. 4-8G-H and appendix 4.4.2.2 for significance values).

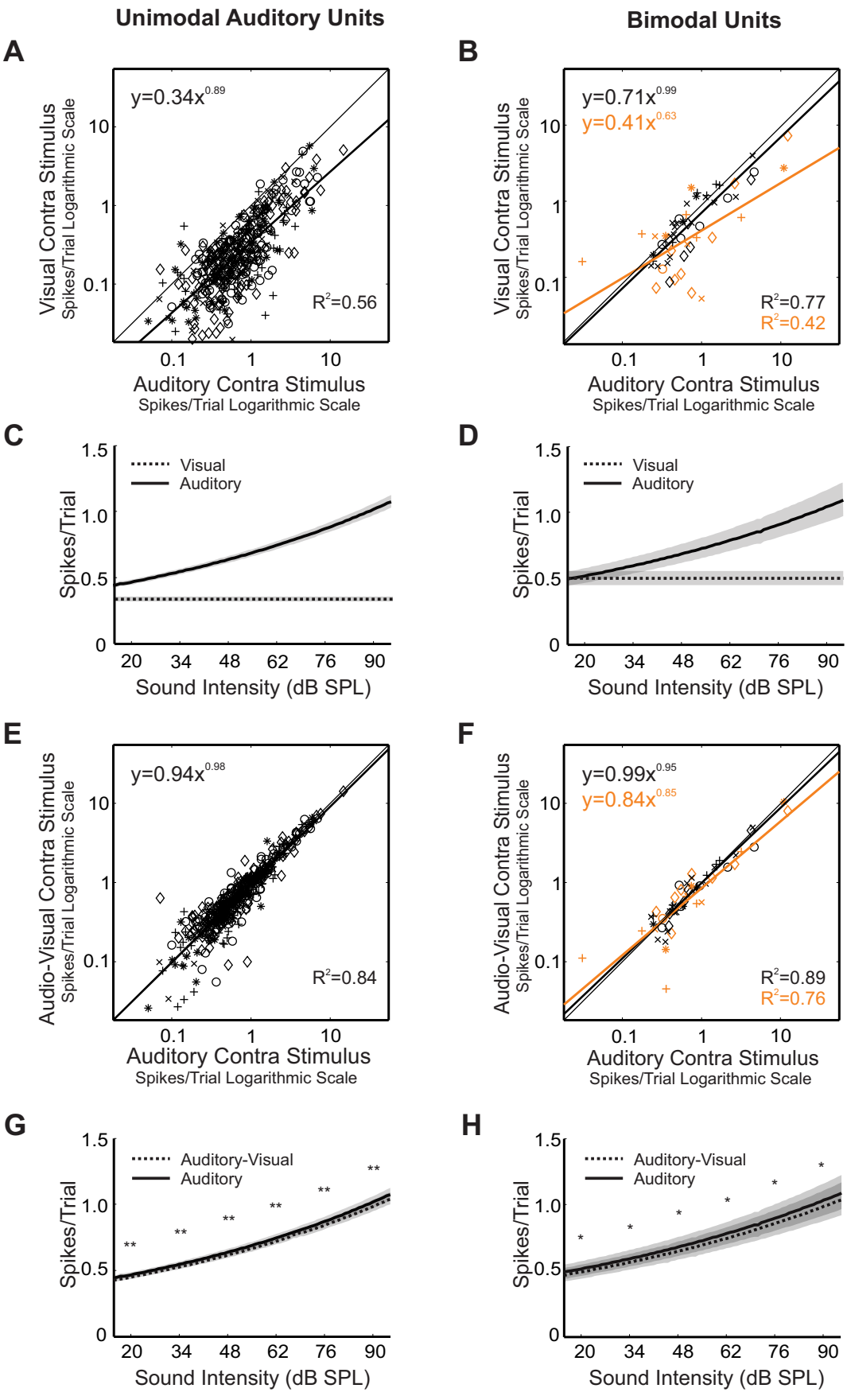

Unimodal Auditory Units
Bimodal Units
A
B
C
D
E
F
G
H
y=0.34x^0.89
R²=0.56
y=0.71x^0.99
y=0.41x^0.63
R²=0.77
R²=0.42
y=0.94x^0.98
R²=0.84
y=0.99x^0.95
y=0.84x^0.85
R²=0.89
R²=0.76
Visual Contra Stimulus
Spikes/Trial Logarithmic Scale
Auditory Contra Stimulus
Audio-Visual Contra Stimulus
Visual
Auditory
Auditory-Visual
Spikes/Trial
Sound Intensity (dB SPL)

Figure 4-8. Effect of stimulus type on spike counts. (A-B) Scatter plots comparing the number of spikes recorded following presentation of contralateral auditory and visual stimuli for units classified as (A) auditory and (B) bimodal (suprathreshold units, black; subthreshold units, orange). Points below the unity line indicate that the recorded activity was higher for auditory than for visual stimuli. A regression line was fitted to each scatter plot, using the linear least squares method, and the corresponding slope and y-intercept are shown along with the R^2 value in each panel. Data from individual animals are represented by different symbols. (C-D) Statistically modelled spike count values shown as a function of sound intensity for auditory (black line) stimuli, together with the modelled spike count value for the fixed intensity visual stimulus. (E-F) Scatter plots comparing the number of spikes evoked by contralateral audio-visual and auditory stimuli for (E) auditory and (F) bimodal units (suprathreshold units, black; subthreshold units, orange). Points below the unity line indicate that units responded more strongly to auditory than to auditory-visual stimuli (G-H) Corresponding statistically modelled spike count values are shown as a function of sound intensity for auditory-visual (dotted line) and auditory (black line) contralateral stimuli below each scatter plot. Significant differences between stimulus types are indicated by asterisks (* $P < 0.05$; ** $P < 0.01$ and *** $P < 0.001$).

4.2.4 Cross-Modal Interactions

The magnitude of the cross-modal interactions exhibited by units was quantified by calculating the percentage difference between the mean number of spikes evoked by auditory-visual and auditory stimuli (Meredith & Stein, 1983, 1986b; King & Palmer, 1985; Stein *et al.*, 2009; Stevenson *et al.*, 2014). A positive response modulation value indicates an enhancement of the auditory response due to the presence of a visual stimulus, whereas a negative value indicates a depression. In the previous section, visual co-stimulation was shown to suppress the average auditory response of the population regardless of whether units were classified as unimodal auditory or bimodal. Therefore, response modulation values were calculated for all units and not only the bimodal (Fig. 4.9A-B). Accordingly, for both unit types, we found that the distribution of response modulation values was skewed towards negative values, indicating that the majority of units exhibit a response depression.

It the SC, a multisensory nucleus, the magnitude of cross-modal interactions has been shown to be inversely related to stimulus effectiveness (Meredith & Stein, 1986; Stein & Meredith, 1993), a phenomenon known as the principle of inverse effectiveness. To investigate whether this principle also applies to multisensory processing in the auditory cortex, units were separated according to whether they exhibited a response depression or enhancement and response modulation values for each sound intensity were calculated (Fig. 4.9C-D). This analysis shows that, even for units classified as unimodal auditory, the maximum response modulation for enhanced responses was 52% and the minimum response modulation for depressed responses was -13%. In comparison, for bimodal units this range was slightly larger, ranging from 57% to -21%. Notably, the largest modulatory effects, regardless of whether firing rate was enhanced or suppressed, were consistently observed at 34 dB SPL, which is close to the average threshold intensity for the auditory stimulus (unimodal auditory units 31.7 ± 0.7 dB SPL; bimodal units 32.8 ± 2.5 dB SPL). In summary, our results show that, like in other species (rhesus monkey: Ghazanfar *et al.*, 2005; Lakatos *et al.*, 2007; Kayser *et al.*, 2009, 2010), the principle of inverse effectiveness also applies to multisensory processing in the ferret auditory cortex. This would seem intuitive from an ethological perspective, given that response amplification becomes redundant when one stimulus is highly salient.

So far, our results have shown that magnitude of cross-modal interactions in the auditory cortex is inversely related to the stimulus effectiveness. Therefore, it seems reasonable to assume that the magnitude of cross-modal interactions might also depend on the strength of the evoked response (Fig. 4-9E-F). For units that exhibit a response enhancement, the results show a significant negative correlation

between the auditory response strength and response modulation values (Fig. 4-9E, Spearman rank correlation: r = -0.46, $P < 10^{-8}$), whereas for depressed units the results show a significant positive correlation (Fig. 4-9F, Spearman rank correlation: r = 0.29, $P < 10^{-8}$). Overall, these results suggest that the magnitude of modulation depends on both the efficacy of the stimulus and the underlying response strength.

The principle of inverse effectiveness has also been shown to predict behavioural performance gains (Stein *et al.*, 1989b; Holmes, 2007). In the present study, auditory cortical neural activity was recorded whilst animals performed a multisensory localisation task, thus enabling us to directly examine the relationship between neural modulation and behavioural performance across stimulus intensities. Our analysis shows that behavioural gains were maximal at the lowest sound intensity tested (20 dB SPL) and declined steadily with increasing stimulus intensity (Fig. 4-9C-D, grey line). On the other hand, the largest modulation of neural responses occurred near the threshold intensity for the auditory stimulus (34 dB SPL) (Fig. 4-9C-D, black lines). Therefore, although the principle of inverse effectiveness is satisfied at both the neural and behavioural levels, their different scales suggest that the behavioural gains observed at the lowest sound intensities were not fully accounted for by changes in the neural firing rate at the level of the auditory cortex alone.

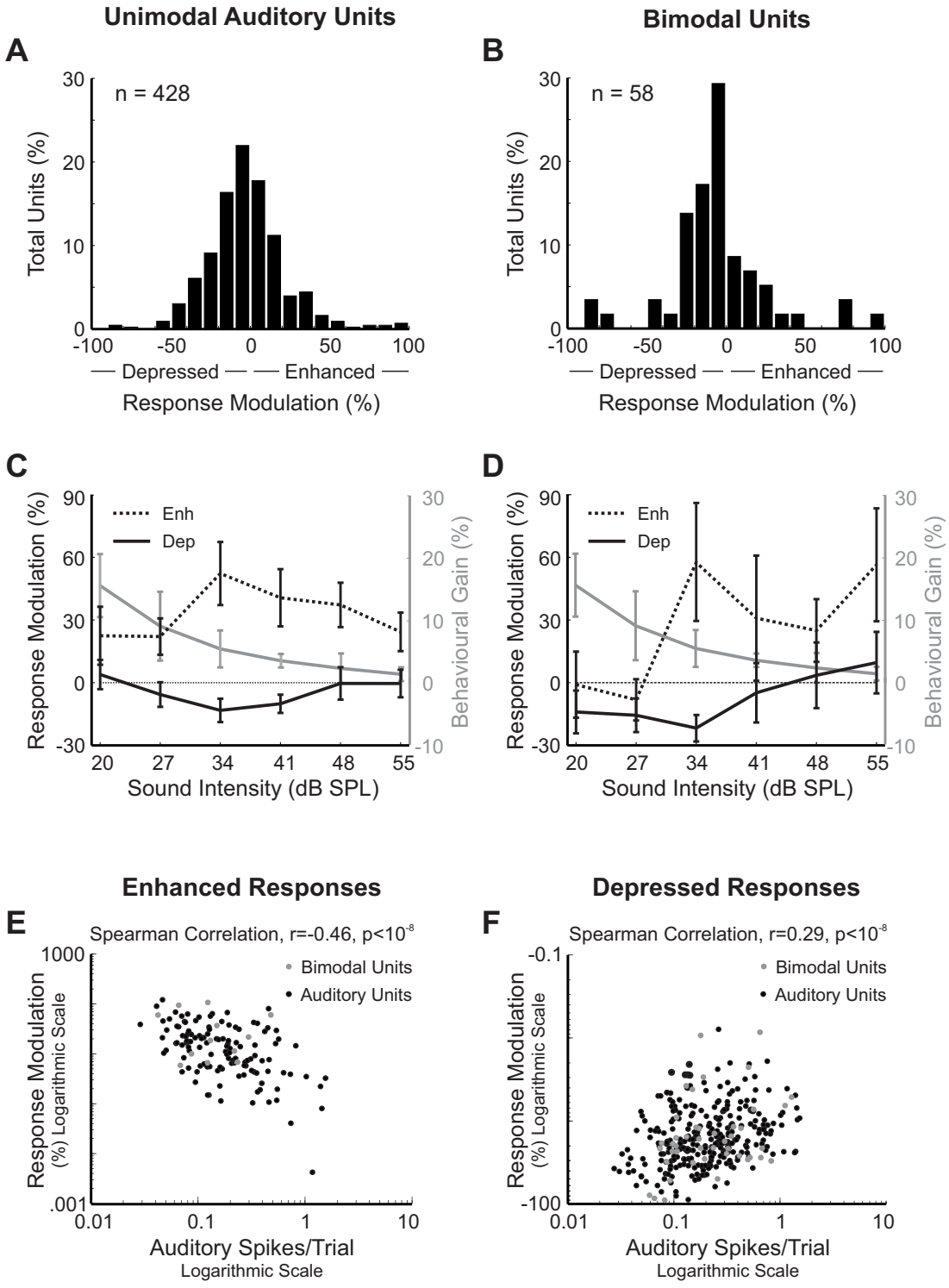


Figure 4-9. Visual modulation of auditory cortical activity. (A-B) Histograms showing the distribution of response modulation values for units classified as unimodal auditory (A) or bimodal (B). Positive and negative values indicate that responses were enhanced and depressed by multisensory stimulation, respectively. (C-D) Response modulation values are shown as a function of sound intensity for units classified as unimodal

auditory (C) or bimodal (D). Units are separated according to whether their responses were enhanced or depressed by multisensory stimulation (enhanced units, black dashed line; depressed units, black solid line). For comparison, the multisensory behavioural gain expressed as a percentage is also shown (grey line). (E-F) Logarithmic scatter plots of auditory response strength (spikes/trial) versus response modulation value. Each dot represents one unit (bimodal units, grey dots; auditory units, black dots). For simplicity, units are shown in separate panels according to whether their responses are enhanced (E) or depressed (F) by multisensory stimulation.

In addition to firing rate, studies of auditory (Brugge *et al.*, 2001; Nelken *et al.*, 2005) and multisensory (Bizley *et al.*, 2007; Ghazanfar & Lemus, 2010; Kayser *et al.*, 2010; Sarko *et al.*, 2013) processing have shown that stimulus information can also be encoded by the temporal features and variability of spike discharge patterns.

In the SC (Rowland *et al.*, 2007) and belt areas of the auditory cortex (Chandrasekaran *et al.*, 2013) response latencies have been shown to be shorter for combined auditory-visual stimuli than for auditory stimuli alone. Therefore, we investigated whether a similar effect exists in the auditory cortex, which would correlate with our behavioural results from the localisation task (see Section 3.2.2) as well as with an extensive body of literature demonstrating that behavioural responses are faster under multisensory conditions. However, an initial inspection of the PSTHs shows that the temporal response dynamics of neurons in the auditory cortex were very similar for auditory-visual and auditory stimuli (Fig 4-10A-B). In order to quantify the effect of stimulus condition on response latency, we fitted a linear mixed effects model to mean latency values, which were derived by calculating the mean latency of all spikes within a 150 ms window following stimulus onset. As expected, mean latencies decreased with increasing sound intensity (see Fig 4-10C and Fig 4-10D for units classified as unimodal auditory and bimodal, respectively). On average, mean latencies were shorter for auditory-visual stimuli than for auditory

stimuli (unimodal auditory units AV vs. A: 61.30 ± 0.48 vs 61.46 ± 0.48 ms; bimodal units AV vs. A: 67.76 ± 1.21 vs 69.72 ± 1.17 ms); however the difference was minimal: 0.16 ms for unimodal auditory units and 1.96 ms for bimodal units. Furthermore, it was not statistically significant for either unit type across sound intensities (range of significance values across sound intensities, auditory units: t_{5135} = -1.45 – 1.88, P = 0.14 – 0.059; bimodal units: t_{5135} = -1.59 – 0.85, P = 0.11 – 0.40). Moreover, these differences are considerably smaller than those values reported previously for the SC (Rowland *et al.*, 2007). In summary, our results suggest that the temporal response dynamics of neurons in early processing areas of the auditory cortex are not affected by multisensory stimulation, which is perhaps not surprising given the temporal disparity of auditory and visual inputs.

Spike discharge pattern variability can be measured by calculating the Fano Factor (FF) (Fano, 1947), which is the ratio of the variance to the mean of spike counts in a given time window. For a Poisson distribution, the most common statistical model used to simulate neural firing (Eden & Kramer, 2010), the mean and variance are equal. Therefore, a Fano Factor of one represents a spike discharge pattern that is as reliable as would be found for a Poisson process (Fano, 1947; Eden & Kramer, 2010). Our analysis found no consitent effect of sound intensity on FF values (data not shown), therefore data were pooled across sound intensities for the analysis. The results suggest that the addition of the visual stimulus had opposing effects on Fano Factor values, depending on the mode of multisensory interaction (i.e. enhancement vs. depression). On average, for both units classified as unimodal auditory or bimodal, neural responses that were enhanced by multisensory stimulation tended to become less reliable (i.e. FF increased, unimodal auditory units

AV vs. A 2.2 ± 0.08 vs. 2.05 ± 0.09; bimodal units AV vs. A 2.13 ± 0.25 vs. 2.08 ± 0.27) under multisensory conditions (Fig. 4.10E-F) but this effect was only statistically significant for units classified as unimodal auditory (t-test, t_5 = -3.70, P = 0.01). In contrast, neural responses that were depressed by multisensory stimulation tended to become more reliable (i.e. FF decreased, unimodal auditory units AV vs. A 2.03 ± 0.07 vs. 2.15 ± 0.07; bimodal units AV vs. A 2.27 ± 0.21 vs. 2.45 ± 0.22) (Fig. 4.10A-B). However, this effect was not statistically significant for either unit type. A similar trend has been reported from single neuron recordings made from the auditory cortex. It has been suggested that these opposing modes of multisensory integration potentially have different functions, whereby enhanced but less reliable responses might aid the detection of near-threshold events. In contrast, suppressed but more reliable responses might be more relevant for stimulus discrimination in complex multisensory environments (Kayser *et al.*, 2010; Sarko *et al.*, 2013).

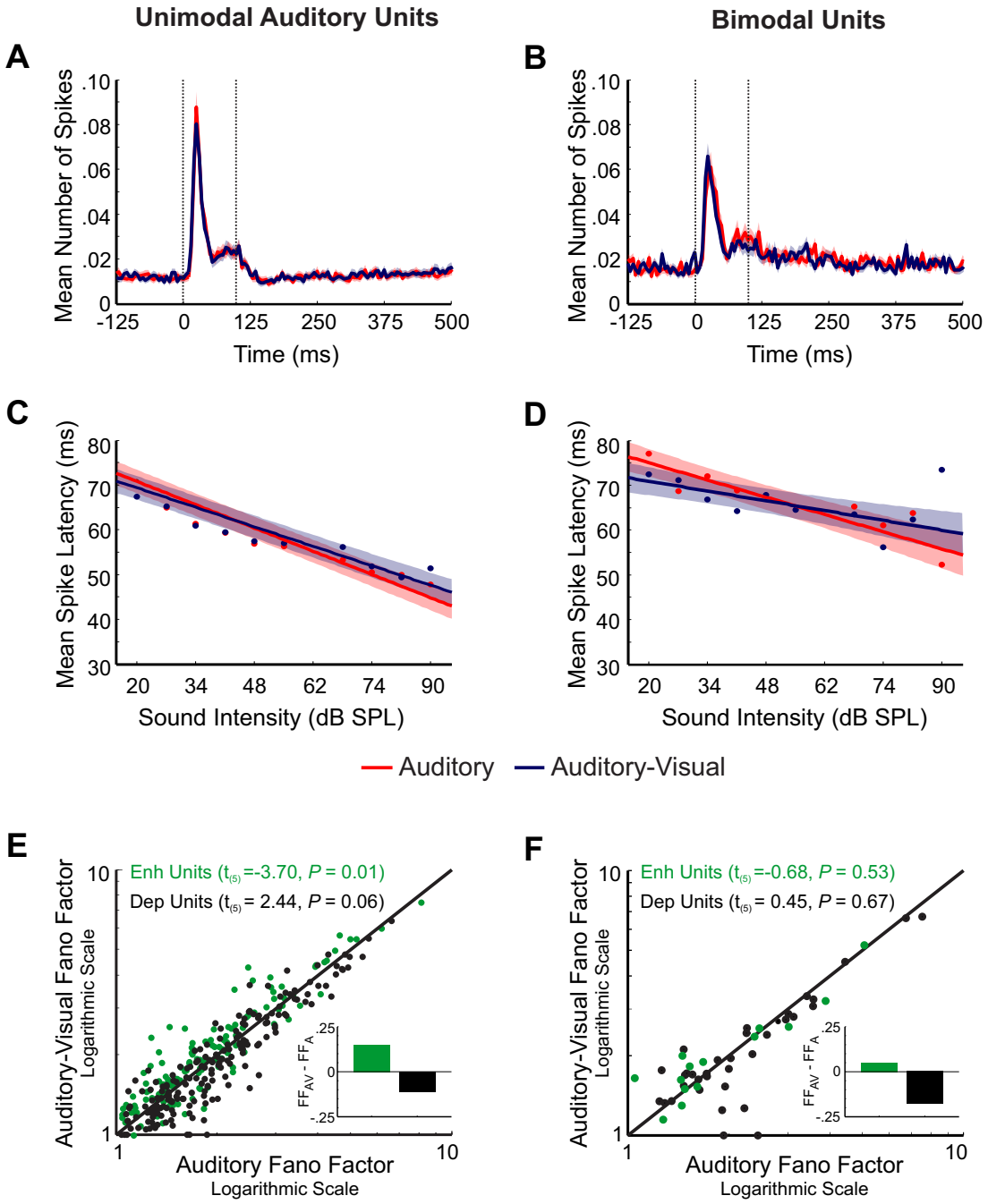


Figure 4-10. Multisensory effects on spike timing and discharge pattern variability. For simplicity, data from units classified as unimodal auditory or bimodal are shown seperately throughout the figure in the left and right panels, respectively. (A-B) Pooled peri stimulus time histograms showing the mean number of spikes evoked by auditory-visual (blue) and auditory (red) stimuli in 5 ms time bins relative to stimulus onset. (C-D) Mean response latency shown as a function of sound intensity for auditory-visual (blue) and auditory stimuli (red). (E-F) Scatterplot comparing Fano Factor values (ratio of variance to mean) on a logarithmic scale for auditory versus auditory-visual stimuli. Each dot represents one unit (green dots, enhanced responses; black dots, depressed responses). Points above the line indicate that responses were less reliable for

the auditory-visual condition, whilst points below the line indicate that responses were more reliable for the auditory-visual condition. (Inset) Bar chart summarising the mean difference between Fano Factor values for auditory-visual and auditory stimuli.

4.2.5 Correlation of Neural Activity and Behavioural Outcomes

Finally, we investigated whether activity in the auditory cortex during the task performance correlated with behavioural outcomes. An inspection of the pooled histograms showing the mean spike rate in 5 ms time bins following stimulus onset shows that the peak rate is higher for recordings made during correct as compared to incorrect approach-to-target responses (Fig. 4.11A-D top panels). Furthermore, this observation was consistent across all stimulus conditions. To explore whether neural activity in the auditory cortex is a significant determinant of a correct response, a generalised linear mixed effects model for Bernoulli-distributed responses was fitted to the approach-to-target response correctness data. The neural parameters of evoked and spontaneous activity (i.e. the number of spikes evoked in a 150 ms window before and after stimulus onset, respectively) were included in the model as fixed effects, in addition to the stimulus parameters of location relative to the unit (i.e. contralateral or ipsilateral), intensity and type. In addition, cell and ferret identities were included in the model as random effects with the former nested within the latter. In keeping with our earlier analysis, the model predicts that the probability of a correct behavioural response depends on the stimulus parameters and increases significantly with increasing sound intensity (log odds ratio = 0.07, $t_{115276} = 33.83$, $P < 0.001$) and also for auditory-visual as compared to auditory stimuli (range of values across stimulus parameters: log odds ratio = 0.44 – 0.59, $t_{115276} = 9.62 – 4.54$,

$P < 0.001$). The modelled relationship between spontaneous (orange lines) and evoked (black lines) activity and the relative behavioural gain across sound intensities is shown in Figure 4.11A-D (bottom panels) and is also summarised by the surface plots in Figure 4.12A-B.

For both auditory and auditory-visual stimuli, the model shows that the probability of a correct approach-to-target response increases with increasing evoked activity in the contralateral hemisphere. This effect was significant across the lower sound intensities (see Fig. 4-11 and appendix 4.4.2.3 for significance values). In this study, we set out to explore whether the auditory cortex acts as a substrate for the visual enhancement of sound localisation behaviour. However, an inspection of the surface plots shows that the probability of a correct response was higher for auditory-visual stimuli than for auditory stimuli at the same evoked activity level. Furthermore, we have shown previously that the addition of the visual stimulus subtly suppressed, rather than enhanced, the evoked activity level. Together, these findings suggest that the visual enhancement of task performance is not attributable to a change in the evoked activity level in early auditory processing areas.

In addition to showing that the probability of a correct response increases with increasing evoked activity level in the contralateral hemisphere, although not statistically siginificant, the model also suggests that the probability of a correct response tended to increase with decreasing spontaneous activity level (log odds ratio = -0.010, $t_{115276} = -0.91$, $P = 0.36$). Taken together, these findings suggest that animals are more likely to make a correct response when the ratio of the evoked to the spontaneous activity (i.e. signal to noise ratio) in the contralateral cortex is high. Consequently, the cross-modal suppression of auditory cortical activity observed

during the task performance might enhance stimulus detection by disproportionally reducing the spontaneous activity level, thus increasing the signal to noise ratio.

In comparison, an inverse relationship was observed in the case of the ipsilateral hemisphere, whereby the probability of a correct response tended to decrease with increasing evoked activity although this effect was not statistically significant (see Appendix 4.4.2.3 for significance values). It has been suggested that sound source azimuth is likely to be encoded by a comparison of the activity between neurons with heterogeneous spatial sensitivity within each hemisphere (Day & Delgutte, 2013; Goodman *et al.*, 2013; Keating *et al.*, 2015). Our findings suggest that sound source location may also be encoded by a comparison of activity across hemispheres. In order to explore this possibility, we fitted a generalised linear model linking the probability of a correct response to the difference in the mean auditory response of the neural populations in the contra- and ipsilateral hemispheres (Fig. 4.13). In support of a hemispheric model, the model output shows that for auditory stimuli, the difference in hemispheric activity is a significant factor in determining the probability of the animal making a correct approach-to-target response (Fig. 4.13A, log odds ratio = 0.018, $t_{10935} = 2.01$, $P = 0.02$), whereby animals are more likely to make a correct response when the activity in the contralateral cortex is high and the ipsilateral cortex is low. In contrast, this effect was not significant for auditory-visual stimuli (Fig. 4.13B, log odds ratio = 0.005, $t_{10935} = 0.46$, $P = 0.64$).

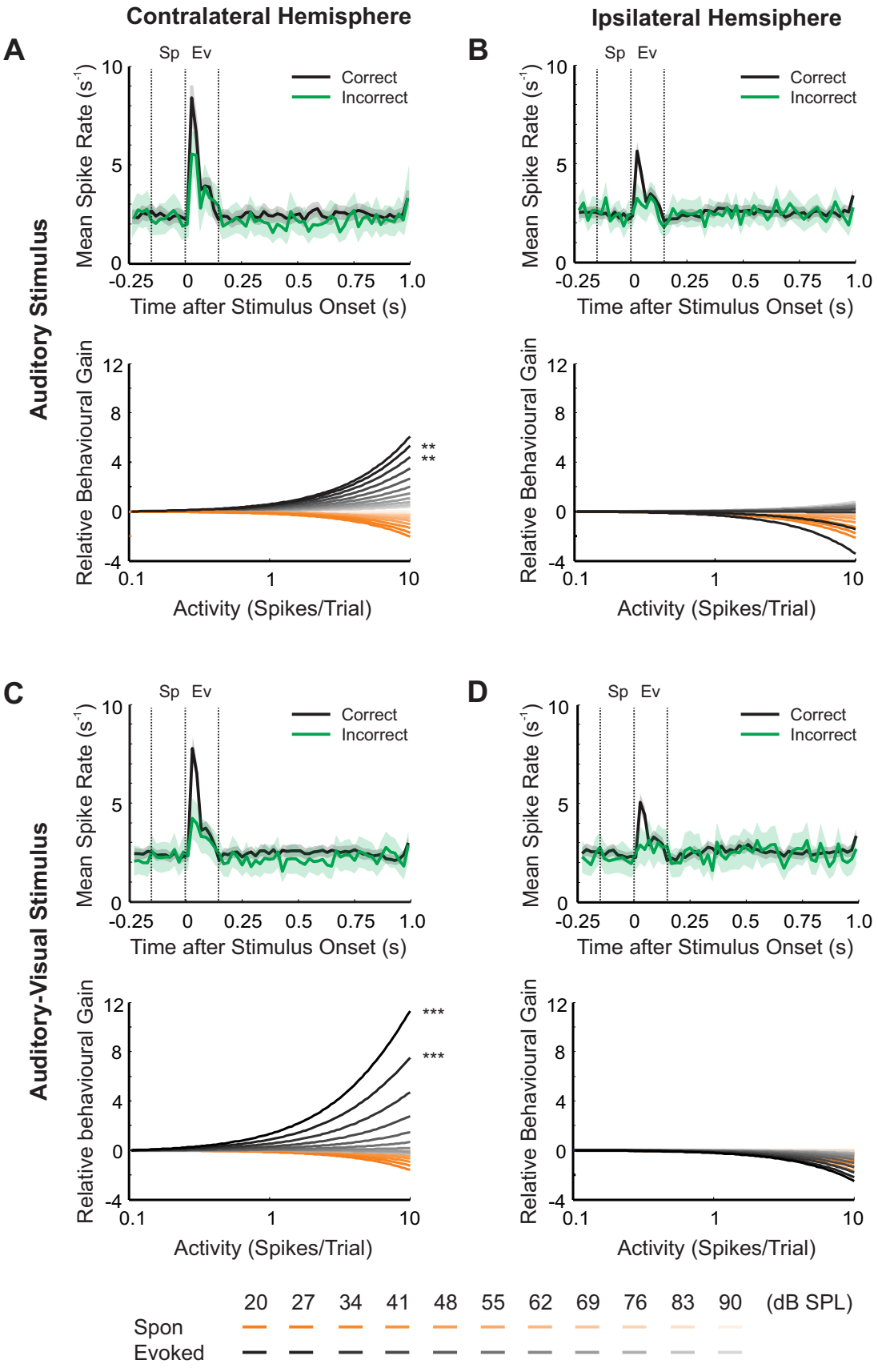


Figure 4-11. (A-D, top panels) Pooled histograms showing the mean spike rate in 5ms

time bins relative to the stimulus onset for correct (black) and incorrect trials (green). For simplicity, data from auditory and bimodal units are pooled throughout the figure. In addition, data are also seperated according to the relative location of the unit to the stimulus (contralateral hemisphere: A, C; ipsilateral hemisphere: B, D). (A-D, bottom panels) Plots showing the modelled relationship between spontaneous (orange) and evoked (black) activity and the relative behavioural gain. A significant change in behavioural gain over evoked activity is indicated by the asterisks and probability value by their number (one: $P < 0.05$; two: $P < 0.01$ and three: $P < 0.001$).

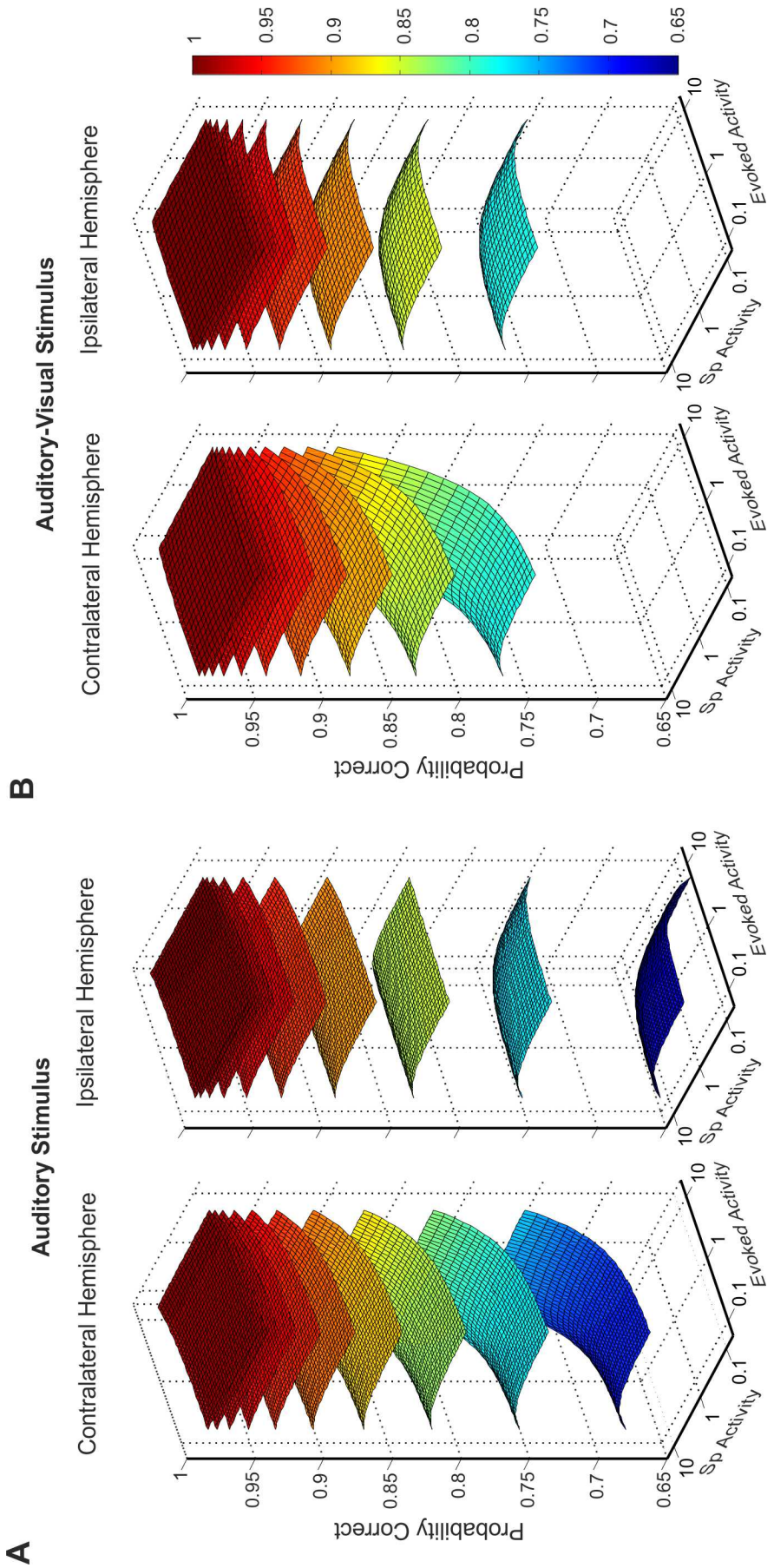


Figure 4-12. (A-D) Surface plots showing the modelled probability of a correct approach-to-target response as a function of spontaneous and evoked activity. The different surfaces within each plot represent the different intensity levels (20 – 90 dB SPL). The model output for each stimulus condition – location combination is shown in the separate panels.

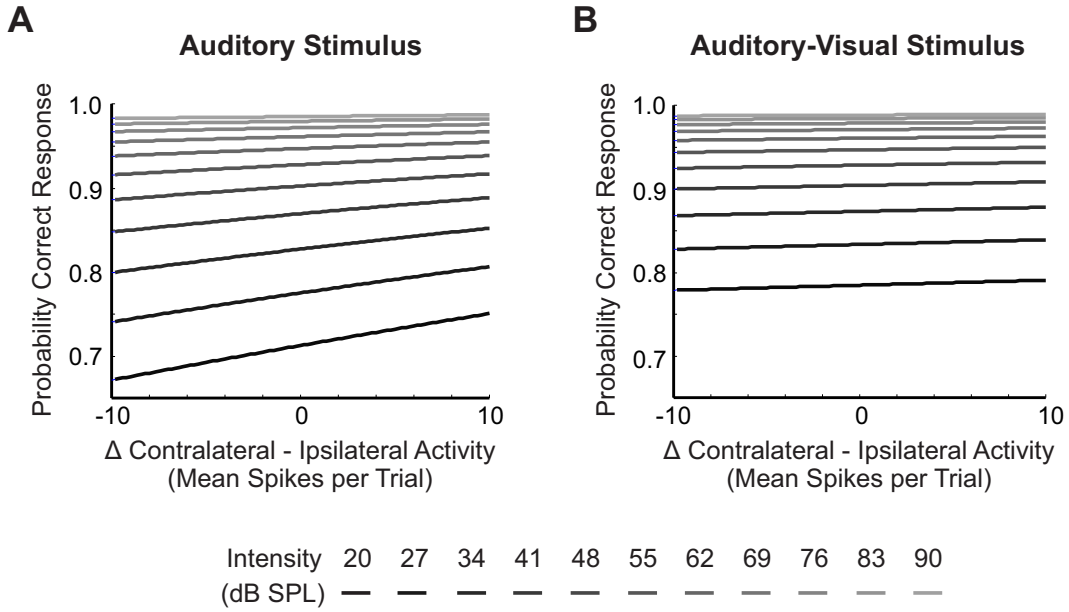


Figure 4-13. (A-B) Plots showing the modelled probability of a correct approach-to-target response versus the difference (Δ) in mean evoked activity between the contra- and ipsilateral hemispheres. Model predictions are shown separately for auditory (A) and auditory-visual (B) stimuli across sound intensities (grey scale).

4.3 Discussion

In opposition to the traditional view that sensory cortices are unisensory, some studies have shown multisensory convergence and interactions in primary and non-primary sensory areas (reviewed by Ghazanfar & Schroeder, 2006). At the level of the auditory cortex, for example, visual inputs have been shown to modulate and even in some instances evoke neural activity (monkeys: Schroeder & Foxe, 2002; Brosch, 2005; Ghazanfar *et al.*, 2005; Kayser *et al.*, 2008; ferrets: Bizley *et al.*, 2007). Recent work published from this laboratory suggests that the auditory cortex may provide a substrate for visual influences on spatial hearing (Bizley & King, 2008). However, in general the perceptual and, ultimately, the behavioural relevance of cross-modal activity in the early auditory cortex remain elusive, since few studies have been performed whilst animals were engaged in a multisensory task.

For the experiments described in this chapter, neural activity in the auditory cortex has been explored whilst animals localised unisensory and multisensory stimuli presented from either the left or the right hemifield. The observed increase in localisation accuracy for auditory-visual targets as compared to either auditory or visual targets in isolation during the recordings provides robust evidence for multisensory integration. In view of the initial independence of the sensory systems at early processing stages, it would seem intuitive that the integration of estimates from two or more senses should yield a more accurate overall estimate. At both the neural and behavioural levels, statistically optimal cue integration models have been shown to predict the output of multisensory integration. For example, when both visual and vestibular cues about the direction of self-motion in the world are

simultaneously available, a subpopulation of multisensory neurons in the dorsal medial superior temporal area of the macaque visual cortex have been shown to combine cues near optimally (Angelaki *et al.*, 2011). Likewise, behavioural precision has also been shown to improve in a statistically optimal manner. Similarly, the results of the present study show that approach-to-target performance approximates that predicted by a maximum likelihood estimator model indicating that, like in humans (Ernst & Banks, 2002; Battaglia *et al.*, 2003; Alais & Burr, 2004), multisensory integration in ferrets occurs in a statistically optimal manner whereby inputs, weighted by their relative reliabilities, are integrated in a linear fashion.

For the neural recordings, a relatively short stimulus duration was chosen (100 ms) to both optimise the experimental design and mitigate several confounding effects. First, because the stimulus duration determined the temporal window that was analysed, this limited potential feedback on auditory cortical activity from other cortical regions, including motor areas and higher associative areas involved in the decision-making process. Second, this ensured that the stimulus was over before the first motor response (i.e. the head orienting movement). Otherwise, the evoked activity may have reflected a relative change in the sound location, in addition to the intended changes in the stimulus characteristics. Third, this ensured that the stimulus was sufficiently close to threshold in order to observe visual enhancement of auditory localisation ability, whilst still above the detection threshold. Although the majority of units exhibited an onset response (mean auditory response latency: 27 ms), some units exhibited an offset response with a similar latency to the onset response from the stimulus offset. This precise temporal pattern of evoked activity determined the temporal window that was considered for the analysis (150 ms).

During the multisensory task, auditory cortical activity was overwhelmingly driven by auditory inputs. In total, auditory inputs drove 94% of units (481/509) and accounted for 96% of evoked activity. Whilst a local place code, comprising a topographic map of interaural level differences and spectral cues, exists in the SC, it has thus far not been demonstrated in the auditory cortex. A number of studies have shown that although most neurons in the auditory cortex are broadly tuned to sounds presented in contralateral space, a small but significant proportion are tuned to sounds presented in ipsilateral space, thus creating left- and right-tuned channels within as well as across hemispheres. Based on this organisation and, in addition, the finding that unilateral cortical lesions produce a sound localisation deficit that is confined to contralateral space (Malhotra & Lomber, 2007), the existence of an opponent channel population code has been proposed, whereby sound location is determined by comparing the activity of left and right tuned channels within the same hemisphere (Stecker *et al.*, 2005; Keating *et al.*, 2015). In the present study, lateralised stimuli were also found typically to evoke lateralised responses, thereby supporting the existence of left- and right-tuned channels across hemispheres. Furthermore, our results suggest that sound source azimuth may also be represented by a hemispheric code, in addition to an opponent channel population code. In support of this hypothesis, we were able to directly link sound localisation behaviour to the right/left balance of activity across hemispheres since our results show that animals were more likely to correctly localise stimuli with increasing evoked activity levels in the contralateral cortex and decreasing levels in the ipsilateral cortex.

An important aspect to consider when exploring audio-visual integration is the prevalence of visual inputs and, in addition, their interaction with auditory

responses. In the SC, it has been estimated that 65% of units are bimodal (Meredith *et al.*, 2012). Furthermore, the integration of multi-modal inputs by SC neurons typically results in a response enhancement, which is commonly in the additive or even supra-additive range (i.e. the combined response is greater than the sum of the unisensory responses) (Meredith & Stein, 1986b). In contrast, a comparatively lower level of integration was observed in the auditory cortex during the localisation task. First, a lower proportion of units were influenced by visual stimulation (16%, 81/509). Although estimates vary between studies, this value is consistent with previous estimates for this species, albeit for anaesthetised animals (Bizley *et al.*, 2007; Meredith & Allman, 2015). Second, whilst a subpopulation of units displayed response enhancement, the majority exhibited response suppression and overall the effect of concurrent visual stimulation on the mean auditory response of the population was suppressive. Third, in terms of the magnitude of response modulation, units exhibited a lower range of integration. Similar results have been reported for the primate auditory cortex (Kayser *et al.*, 2008) and mouse visual cortex (Meijer *et al.*, 2017) suggesting that a predominantly low-level modulatory form of multisensory integration is a common characteristic of early sensory cortices, which in the context of their functional roles would seem appropriate.

The temporal profile of multisensory integration has been proposed to take different forms, depending on the timing of inputs (Rowland and Stein, 2008). In the SC (Rowland *et al.*, 2007) and belt areas of the auditory cortex (Chandrasekaran *et al.*, 2013), the simultaneous arrival of inputs has been shown to result in real-time integration. In this model, the arrival of simultaneous inputs enables the response threshold to be reached earlier and, consequently, the shortening of response

latencies. In contrast, in the present study it is clear from the response latency data that a significant disparity exists between the timing of auditory and visual inputs (27 vs. 57 ms). A similar temporal disparity has been described for the macaque auditory association cortex (Schroeder & Foxe, 2002, 2005) and ferret auditory cortex (Bizley *et al.*, 2007). Moreover, we found that the largest cross-modal modulatory effect on spike rate occurred ~90 ms after stimulus onset. Overall, our results fit well with a delayed integration model, whereby the existence of a temporal disparity between inputs by default places a lower bound on the timing for neural integration. Consequently, integration can only occur later in the response and, therefore, has no effect on the response latency. Although delayed integration limits the temporal window over which integration can occur, it would nevertheless confer auditory inputs preferential access to local cellular excitability and, in addition, the potential ability to modulate the processing of the later inputs (Schroeder & Foxe, 2002).

In some units, suprathreshold visual inputs were able to evoke activity **(**16%, 81/509**)**. However, this visual-evoked activity typically lacked a clear temporal onset. Furthermore, it was substantially weaker than auditory-evoked activity, in terms of the evoked activity expressed as a percentage increase of the spontaneous activity (V vs. A: 42% vs. 113%). Therefore, it is perhaps not surprising that, unlike for auditory-evoked activity, visual-evoked activity was not significantly related to the probability of a correct localisation response. Whilst it is possible that a more naturalistic stimulus might have increased the prevalence of visually responsive units and/or evoked stronger responses, our behavioural results from the localisation task confirm that the LED used was sufficiently salient to evoke multisensory integration. In addition, auditory-visual interactions in the rhesus monkey auditory cortex have

been shown to be comparable for both naturalistic scenes and simple noise bursts and flash stimuli (Kayser *et al.*, 2008). Whereas only a small proportion of units were driven by suprathreshold visual inputs, a greater proportion of units were modulated by subthreshold visual inputs. Furthermore, a modulatory effect was even apparent on the mean auditory response of the population of units classified as unimodal auditory. The latter finding highlights the need to study multisensory integration at the population level, a conclusion also drawn from studies of multisensory processing in the extrastriate visual cortex (Allman *et al.*, 2009) and auditory cortex (Meredith & Allman, 2015). In the ferret, the relative proportion of neurons that receive subthreshold inputs that modulate activity versus suprathreshold inputs that drive activity has been shown to be inversely related between primary and high-level auditory cortical areas (Bizley & King, 2009). The reversal of this relationship mirrors an increase in the density of projections from non-lemniscal thalamic nuclei (Winer & Lee, 2007) and other sensory cortices (Bizley *et al.*, 2007) and suggests a progression from areas with a primary unisensory function to those more involved in merging independent inputs from different sensory modalities. This observation also fits with our histological analysis showing that the majority of recording sites were located in early auditory processing areas (see Appendix A).

Overall, concurrent visual stimulation suppressed evoked activity by 3.2% across the entire population of units sampled in the auditory cortex. A similar suppressive effect has been described for anaesthetised ferrets (Bizley *et al.*, 2007; Meredith & Allman, 2015) and rhesus monkeys (Kayser *et al.*, 2008). An important aspect to consider is the origin of this suppression and its neural basis. However, deciphering the specific underlying circuits is not a straight forward task, since it is

becoming clear that multiple circuits (e.g. thalamo-cortical and cortico-cortical) likely exist for mediating the influence of one modality on another (Iurilli *et al.*, 2012; Song *et al.*, 2017). For example, in rodents, activation of one primary sensory cortex has recently been shown to inhibit neighbouring primary sensory cortices (Iurilli *et al.*, 2012; Stehberg *et al.*, 2014). In terms of the neural circuitry underlying this effect, activation of auditory cortical neurons has been shown to drive local GABAergic inhibition of supragranular pyramidal cells in primary visual and somatosensory cortices via direct cortico-cortical connections and thereby reduce cortical responsiveness (Iurilli *et al.*, 2012). Anatomical studies in the ferret also support the existence of direct projections between primary auditory and visual cortices (Bizley *et al.*, 2007). Consequently, it is conceivable that a similar circuit could contribute to the suppressive effect observed in the present study, whereby activation of the visual cortex could recruit local inhibitory circuits in the auditory cortex via direct connections between the respective sensory cortices. Furthermore, this pathway could also account for the sparse visual-evoked activity observed in the present study since, in the mouse, inhibition was generated by the sound-driven excitation of a few infragranular cells in the visual cortex (Iurilli *et al.*, 2012).

From an evolutional perspective, it would seem reasonable to assume that multisensory convergence in the auditory cortex should have an advantageous effect on its primary role in hearing. An example of this is provided by speech perception. Orofacial movements during speech production provide temporally correlated cues (Chandrasekaran *et al.*, 2013) which, when combined with acoustic signals, improve the detection and comprehension of speech, particularly if auditory signals are degraded by the presence of background sounds (Sumby & Pollack, 1954; Grant &

Seitz, 2000). A growing body of evidence suggests that auditory and visual speech signals interact as early as the primary auditory cortex (Schroeder *et al.*, 2008; Ghazanfar, 2009; Okada *et al.*, 2013). In the present study, the simultaneous recording of auditory cortical activity whilst animals performed the auditory-visual localisation task enabled us to examine the functional relevance of multisensory interactions in the auditory cortex to a different aspect of auditory perception, in the form of spatial hearing. Overall, our results suggest that early auditory cortical processing areas do not act as a substrate for visual influences on spatial hearing. First, behavioural gains peaked at 20 dB SPL, the lowest sound intensity tested, whilst the greatest modulation of evoked activity occurred at 34 dB SPL, close to the threshold intensity of units for the auditory stimulus. Second, the probability of a correct localisation response increased as evoked activity levels in the contralateral cortex increased, whereas visual co-stimulation suppressed the mean auditory response of the population. Third, inspection of the surface plots (Fig. 4-12) shows that at the same evoked activity level, the probability of a correct localisation response was higher for auditory-visual stimuli than for auditory stimuli, illustrating the absence of a simple relationship between the effects of visual cues on localisation performance and auditory cortical activity. This conclusion is perhaps not surprising given the ubiquitous nature of multisensory interactions and, therefore, the likelihood that other cortical areas, including higher-level auditory areas and/or the visual cortex, contributed to neural integration during the localisation task performance.

Therefore, the question remains as to the function of early multisensory interactions. A range of functions have been suggested for inter-areal inhibition, including the potential to degrade distracting sensory processing in neighbouring

cortices, which in turn might enable sensory cortices to compete for the activation of higher cortical areas. Alternatively, they may subserve competitive attentional interactions between sensory modalities (Iurilli *et al.*, 2012). The present study also suggests a potential role for multisensory interactions in the early auditory cortex in facilitating detection. Our results suggest that correct behavioural responses tended to be associated with lower levels of spontaneous activity and higher levels of evoked activity (i.e. a higher signal to noise ratio). Therefore, by reducing the population spontaneous activity, cross-modal suppression of auditory cortical activity may increase the signal to noise ratio and thereby enhance detection. In support of this hypothesis, behavioural studies have shown that irrelevant tactile (Gillmeister & Eimer, 2007) and visual (Odgaard *et al.*, 2004) stimuli not only enhance the detection of faint tones but also increase their perceived loudness. Likewise, irrelevant auditory stimuli have also been shown to enhance the detection and increase the perceived brightness of visual stimuli (Stein *et al.*, 1996; Odgaard *et al.*, 2003).

4.4 Appendix

4.4.1 Supplementary Behavioural Data for the Localisation Task

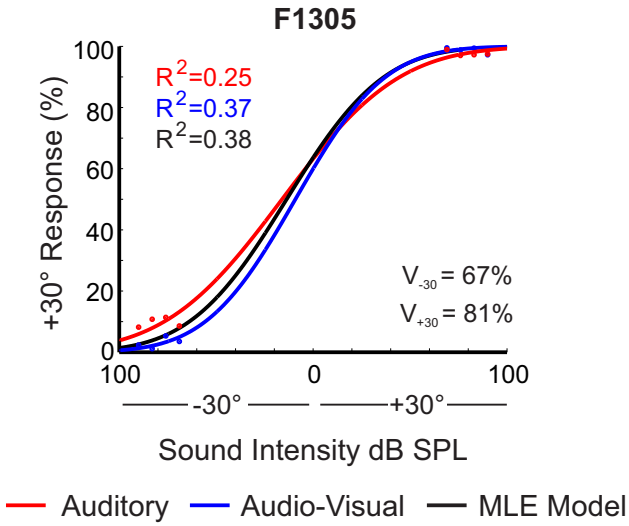


Figure 4-14. Approach-to-target localisation performance for ferret F1305 during the neural recordings. Psychometric curves, fitted with a cumulative Gaussian distribution function, show the percentage of responses to the right-sided spout (+30°) as a function of both stimulus intensity and location for auditory (red) and audio-visual (blue) stimuli. The R^2 values for each curve are shown in the corresponding colour. Modelled audio-visual psychometric curves (black) derived using maximum likelihood estimation are also shown, together with a corresponding R^2 values for the model fit to the measured audio-visual percentage values. Finally, approach-to-target percent correct scores for left (V_{-30}) and right (V_{+30}) sided visual stimuli are also shown.

	Intensity Value 50%			Intensity Value 75%			Slope Magnitude		
	A	AV	MLE	A	AV	MLE	A	AV	MLE
F1305	-16.0	-9.2	-13.7	24.0	23.0	19.1	0.8	1.1	1.0

Table 4-4. Psychometric curve parameters. Psychometric curves were fitted to relate the percentage of trials on which the animal made a right-sided (+30°) approach-to-target response to stimulus intensity and location. Intensity values corresponding to the 50% and 75% probability points are shown along with the slopes of curves fitted to auditory (A), auditory-visual (AV) and modelled (MLE) data.

4.4.2 Statistical Analyses

4.4.2.1 Approach-to-Target Performance

To investigate the effect of stimulus modality, intensity and location on approach-to-target response correctness a generalised linear mixed effects model for Bernoulli distributed responses was fitted to the data. The R code used and model output reads:

```
library(MASS)
mod.all <- glmmPQL(Correctness ~ Stimulus_Type*Signal*I(Sound_Level
- 20), random = ~1|Ferret, family = binomial, data = Ratt)
```

```
                              Value Std.Error    DF t-value p-value
(Intercept)                   1.339     0.132 22805  10.123   0.000
Stimulus_TypeA               -0.506     0.101 22805  -5.017   0.000
Stimulus_TypeV               -0.764     0.097 22805  -7.839   0.000
SignalR                       0.148     0.113 22805   1.306   0.192
I(Sound_Level - 20)           0.034     0.004 22805   8.163   0.000
...
```

The model output shows, for example, that the log-odds corresponding to the auditory modality at -30° and 20 dB SPL is 1.3994 - 0.5062. From this value it is possible to derive the probability (P_A) of a correct response to the auditory target:

$$\log\left(\frac{P_A}{1-P_A}\right) = 1.3994 - 0.5062 = 0.8332$$

$$\frac{P_A}{1-P_A} = \exp(0.8332) = 2.3006$$

$$P_A = 2.3006(1-P_A)$$

$P_A + 2.3006P_A = 2.3006$

$P_A = 0.697$

Consecutive re-levelling of the model for stimulus location and adjustment of the intercept for stimulus intensity yielded the results shown in Table 4-5.

Sound Intensity (dB SPL)	Probability AV	Probability A	*P*-value (*t*-value) AV vs. A
Stimulus Location -30°			
20	0.792	0.697	<0.001 (-5.017)
27	0.829	0.753	<0.001 (-6.121)
34	0.860	0.802	<0.001 (-6.502)
41	0.886	0.843	<0.001 (-5.017)
48	0.907	0.877	0.001 (-3.284)
55	0.925	0.904	0.033 (-2.314)
Stimulus Location +30°			
20	0.816	0.702	<0.001 (-5.782)
27	0.865	0.797	<0.001 (-6.110)
34	0.903	0.867	<0.001(-4.620)
41	0.931	0.916	0.029 (-2.189)
48	0.952	0.948	0.537 (-0.618)
55	0.966	0.968	0.784 (0.784)

Table 4-5. Results of the generalised linear model fitted to the approach-to-target response correctness data. Modelled probabilities of a correct approach-to-target response to auditory-visual (AV) and auditory (A) stimulus targets are shown for each stimulus location-intensity combination, together with the P-values and t-values (in brackets) for the following comparisons: auditory vs. auditory-visual (A vs. AV).

To check the fit of the generalised linear model, a half normal plot was generated (Fig. 4-15). Ordered absolute values of the Pearson residuals were plotted against the expected values for a binomial distribution and, in addition, a simulated envelope was generated, which corresponds to a 95% confidence interval. The model fit is good if the plotted points fall within the boundaries of the simulated envelope.

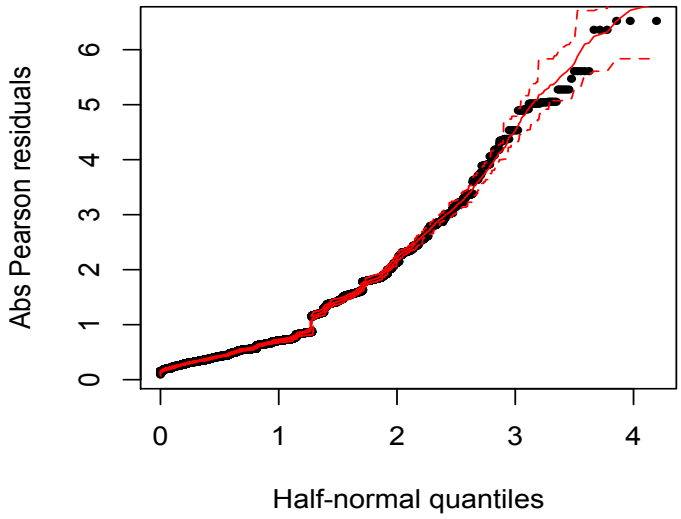


Figure 4-15. Half-normal plot of the Pearson residuals from the generalised linear model fitted to approach-to-target data. A simulated envelope (dashed lines) is also plotted together with the means of the simulated values (solid line). The majority of points fall within the simulated envelope, which indicates a good model fit.

4.4.2.2 The Effect of Stimulus Parameters on the Number of Spikes Evoked Per Trial

A negative binomial generalised linear mixed effects model was fitted to the spike count data obtained during the localisation task. Mixed effects modelling enabled the contribution of fixed effects (i.e. stimulus type, stimulus intensity, stimulus location, cell type and spontaneous count) on the number of spikes evoked per trial (150 ms

window) to be established, while controlling for random effects (i.e. cell identity nested within ferret identity). Inspection of the data showed that a negative binomial rather than a Poisson model was required since the data was over-dispersed (i.e. the variance exceeded the mean). The model was fitted in R using the following code:

```
library(MASS)
mod1 <- glmmPQL(formula(paste("Res_Activity ~ Bgd_Activity +
Cell_Type*Location*Stimilus_Type*I(Intensity - ",i,")")), random = ~
1|Animal/Cell, family = neg.bin(theta = 0.79898), data = Data)
```

	Value	Std.Error	DF	t-value	pvalue
(Intercept)	-1.0814724	0.06247543	121560	-17.31036	0.0000
Bgd_Activity	0.1574399	0.00380178	121560	41.41214	0.0000
Cell_Type2	-0.3089555	0.19534498	502	-1.58159	0.1144
Cell_Type3	0.2476979	0.12355546	502	2.00475	0.0455
LocCL	0.3597835	0.02317402	121560	15.52529	0.0000
Stim_TypeAV	-0.0052529	0.02418472	121560	-0.21720	0.8281
I(Int - 20)	0.0113885	0.00062819	121560	18.12904	0.0000
Cell_Type2:LocCL	-0.1436333	0.10601614	121560	-1.35483	0.1755
Cell_Type3:LocCL	-0.2766683	0.05901075	121560	-4.68844	0.0000
Cell_Type2:Stim_TypeAV	-0.0083388	0.11174444	121560	-0.07462	0.9405
Cell_Type3:Stim_TypeAV	-0.0616438	0.06095211	121560	-1.01135	0.3119
LocCL:Stim_TypeAV	-0.0179180	0.03310730	121560	-0.54121	0.5884
Cell_Type2:I(Int - 20)	-0.0142892	0.00317962	121560	-4.49399	0.0000
Cell_Type3:I(Int - 20)	-0.0057927	0.00178857	121560	-3.23872	0.0012
LocCL:I(Int - 20)	-0.0007689	0.00072482	121560	-1.06075	0.2888

...

The model output shows that for unimodal auditory units (Cell_Type1) contralateral auditory stimuli evoked significantly (*t*-value = 15.52, *P*-value < 0.001) more spikes per trial than ipsilateral auditory stimuli at 20 dB SPL. The ratio of contralateral to ipsilateral spikes per trial was obtained from the model output above as follows:

$\log \mu_{IL} = -1.081 + 0.011(I - 20)$

$\log \mu_{CL} = (-1.081 + 0.360) + (0.011 - 0.001)(I - 20)$

$\log \mu_{CL} - \log \mu_{IL} = 0.360 - 0.001(I - 20)$

and

$\frac{\mu_{CL}}{\mu_{IL}} = \exp\,(0.360 - 0.001(I - 20))$

therefore, since I = 20

$\log \mu_{CL} - \log \mu_{IL} = 0.360$, and $\frac{\mu_{CL}}{\mu_{IL}} = \exp(0.360) = 1.433$

Consecutive re-levelling of the model for stimulus location and adjustment of the intercept for stimulus intensity yielded the results shown in Table 4-6.

Intensity (dB SPL)	**Δ log Count**	$\frac{\textbf{Count}_{\textbf{CL}}}{\textbf{Count}_{\textbf{IL}}}$	**t-value**	**p-value**
Unimodal auditory units: contralateral vs. ipsilateral auditory stimulus				
20	0.360	1.433	15.525	<0.001
34	0.349	1.417	21.516	<0.001
48	0.338	1.403	24.213	<0.001
62	0.327	1.388	17.943	<0.001
76	0.317	1.373	12.172	<0.001
90	0.306	1.360	8.735	<0.001
Unimodal auditory units: contralateral vs. ipsilateral auditory-visual stimulus				
20	0.342	1.408	14.436	<0.001
34	0.338	1.403	20.419	<0.001
48	0.335	1.398	23.458	<0.001
62	0.331	1.393	17.757	<0.001
76	0.328	1.388	12.342	<0.001
90	0.324	1.383	12.324	<0.001

Bimodal units: contralateral vs. ipsilateral auditory stimulus				
20	0.083	1.086	1.532	0.126
34	0.178	1.195	5.099	<0.001
48	0.274	1.315	7.601	<0.001
62	0.369	1.447	6.563	<0.001
76	0.464	1.591	5.623	<0.001
90	0.560	1.750	5.053	<0.001
Bimodal units: contralateral vs. ipsilateral auditory-visual stimulus				
20	0.178	1.195	3.157	<0.001
34	0.246	1.278	9.762	<0.001
48	0.314	1.368	8.700	<0.001
62	0.381	1.465	6.845	<0.001
76	0.449	1.567	5.473	<0.001
90	0.517	1.678	4.683	<0.001

Table 4-6. Results of the generalised linear mixed effects model fitted to investigate the effect of stimulus location on spike counts. The ratio of spikes evoked by contralateral as compared to ipsilateral stimuli is shown separately for auditory and bimodal units and for each stimulus intensity and modality. Corresponding *t*-values and *P*-values are also shown for the following comparison: ipsilateral versus contralateral (CL vs. IL). A significant intensity-location interaction was found for unimodal auditory (*t*-value = 13.503, *P*-value = 0.000) and bimodal (*t*-value = 6.869, *P*-value = 0.000) units indicating that the rate of change of the spike count ratio increases significantly with intensity.

```
mod1 <- glmmPQL(formula(paste("Res_activity ~ Bgd_Activity +
Cell_Type*(Stimilus_Type + I(Intensity - ",i,")*Location)")), random
= ~ 1|Animal/Cell, family = neg.bin(theta = 0.79898), data = Data)
```

Consecutive re-levelling of the model for stimulus modality and adjustment of the intercept for stimulus intensity yielded the results shown in Table 4-7.

Intensity (dB SPL)	**Δ log Count**	$\frac{\mathbf{Count}_V}{\mathbf{Count}_A}$	**t-value**	**p-value**
Unimodal auditory units: auditory vs. visual contralateral stimulus				

20 - 90	-0.304	0.738	-15.102	<0.001
Bimodal units: auditory vs. visual contralateral stimulus				
20 - 90	-0.038	0.963	-0.871	0.384
Intensity (dB SPL)	**Δ log Count**	$\frac{\mathbf{Count_{AV}}}{\mathbf{Count_{A}}}$	**t-value**	**p-value**
Unimodal auditory units: auditory vs. auditory-visual contralateral stimulus				
20 - 90	-0.028	0.972	-2.745	0.006
Bimodal units: auditory vs. auditory-visual contralateral stimulus				
20 – 90	-0.047	0.954	-1.975	0.048

Table 4-7. Results of the generalised linear mixed effects model fitted to investigate the effect of stimulus type on spike count. The ratio of spikes evoked by auditory (A) as compared to visual (V) stimuli and the ratio of spikes evoked by auditory-visual (AV) as compared to auditory (A) stimuli are shown, together with the corresponding *t*- and *P*-values, for auditory and bimodal units. Note that the coefficients are constant across intensities since a stimulus type-intensity interaction was not included in the model.

4.4.2.3 The Effects of Background and Evoked Activity on Approach-to-Target Performance

To investigate the effects of spontaneous and evoked activity (i.e. the number of spikes evoked per trial) on approach-to-target response correctness for different stimulus intensities, types (auditory-visual and auditory) and locations (-30° and +30°), a generalised linear mixed effects model for Bernoulli distributed responses was fitted to the data. The R code reads:

```
modB3.me <- glmmPQL(Correctness ~ Bgd_Activity + Res_Activity +
Loc*Stimilus_Type*Intensity, random = ~ 1|Animal/Cell, family =
binomial, data = Rec)
```

Consecutive re-levelling for stimulus modality and location and adjustment of the

intercept for stimulus intensity yielded the results shown in Table 4-8.

	Evoked Activity			Stimulus Type			Stimulus Location		
dBSPL	Value	*t*-value	*p*-value	Value	*t*-value	*p*-value	Value	*t*-value	*p*-value
	Contralateral Auditory Stimulus								
20	0.031	1.78	0.07	0.44	9.62	<0.001	-0.11	-2.69	0.007
34	0.036	2.98	0.003	0.47	14.4	<0.001	0.09	3.10	0.002
48	0.041	2.62	0.008	0.50	11.12	<0.001	0.30	6.82	<0.001
62	0.047	1.89	0.058	0.53	7.52	<0.001	0.50	7.28	<0.001
76	0.052	1.48	0.138	0.56	5.62	<0.001	0.70	7.26	<0.001
90	0.058	1.25	0.213	0.59	4.54	<0.001	0.91	7.19	<0.001
	Ipsilateral Auditory Stimulus								
20	-0.015	-0.79	0.43	0.56	12.42	<0.001	0.11	2.69	0.007
34	-0.001	0.043	0.97	0.53	16.17	<0.001	-0.09	-3.10	0.002
48	0.014	0.68	0.50	0.51	9.87	<0.001	-0.30	-6.83	<0.001
62	0.029	0.87	0.38	0.49	5.92	<0.001	-0.50	-7.28	<0.001
76	0.044	0.93	0.35	0.46	4.01	<0.001	-0.70	-7.26	<0.001
90	0.056	1.25	0.21	0.59	4.54	<0.001	0.91	7.19	<0.001
	Ipsilateral Auditory-Visual Stimulus								
20	-0.014	-0.65	0.52	-0.56	-12.42	<0.001	-0.014	0.028	0.98
34	-0.019	-1.23	0.22	-0.53	-16.17	<0.001	-0.15	-4.25	<0.001
48	-0.024	-1.08	0.28	-0.51	-9.87	<0.001	-0.30	-5.69	<0.001
62	-0.029	-0.82	0.41	-0.49	-5.92	<0.001	-0.45	-5.40	<0.001
76	-0.034	-0.68	0.49	-0.46	-4.01	<0.001	-6.01	-5.11	<0.001
90	-0.039	-0.60	0.54	-0.44	-2.94	0.003	-0.76	-4.92	<0.001
	Contralateral Auditory-Visual Stimulus								
20	0.09	3.93	0.0001	-0.44	-9.62	<0.001	0.001	0.028	0.98
34	0.66	3.88	0.0001	-0.47	-14.43	<0.001	0.15	4.25	<0.001
48	0.04	1.83	0.06	-0.50	-11.13	<0.001	0.30	5.69	<0.001
62	0.008	0.27	0.78	-0.53	-7.52	<0.001	0.45	5.40	<0.001

76	-0.019	-0.46	0.64	-0.56	-5.62	<0.001	0.60	5.11	<0.001
90	-0.05	-0.86	0.39	-0.59	-4.54	<0.001	0.76	4.92	<0.001

Table 4-8. Results of the generalised linear model fitted to the approach-to-target response correctness data. Statistical values for each of the fixed effects are shown for each stimulus intensity-location-condition combination.

To check the fit of the generalised linear mixed effects model a half normal plot was generated (Fig. 4-16). Ordered absolute values of the Pearson residuals were plotted against the expected values for a binomial distribution and, in addition, a simulated envelope was generated, which corresponds to a 95% confidence interval. The model fit is good if the plotted points fall within the boundaries of the simulated envelope.

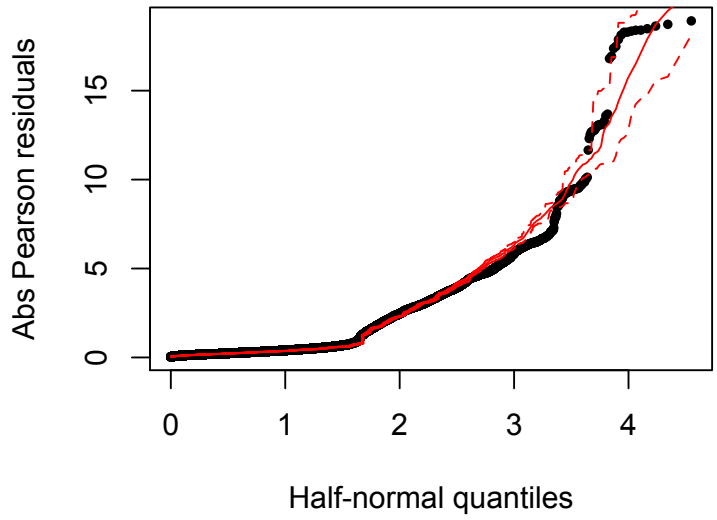


Figure 4-16. Half-normal plot of the Pearson residuals from the generalised linear model fitted to approach-to-target data. A simulated envelope (dashed lines) is also plotted together with the means of the simulated values (solid line). The majority of points fall within the simulated envelope, which indicates a good model fit.

5 CATEGORISATION BEHAVIOUR BASED ON MULTISENSORY BINDING

5.1 Introduction

The integration of sensory inputs from different modalities not only confers behavioural advantages, including enhanced localisation accuracy and reduced reaction times as explored in chapters 3 and 4, but is also essential for scene analysis. Our tendency to merge auditory-visual inputs is illustrated by the ventriloquism illusion (Thomas, 1941; Witkin *et al.*, 1952), whereby auditory stimuli are mis-localised towards visual stimuli provided they are presented at the same time. Likewise, the tendency to merge auditory-visual speech cues is illustrated by the McGurk illusion (McGurk & MacDonald, 1976), whereby pairing a voice articulating one syllable with a face articulating a different syllable can result in listeners perceiving a third sound that represents a fusion of those syllables.

However, how auditory and visual information are integrated to form a coherent perceptual object, despite their fundamental differences in energy forms, transduction mechanisms and initial neural coding strategies, remains to be determined.

Psychophysical studies, including those of sensory illusions, have deciphered a variety of factors that determine whether auditory and visual stimuli are integrated including: low-level physical factors, such as spatio-temporal co-incidence (reviewed by Stein, 2012), as well as high-level cognitive factors, such as semantic congruency (Laurienti *et al.*, 2004; Molholm *et al.*, 2004), assumption of unity (Welch & Warren, 1980; Warren *et al.*, 1981; Vatakis & Spence, 2007) and the interplay between intra- and crossmodal binding (Spence, 2007). Typically, studies have been performed in humans and have used subjective measures of perceptual fusion such as judging the point of spatial alignment for a horizontally moving visual stimulus that crossed the position of a stationary sound source (Lewald *et al.*, 2001) or the simultaneity of a light flash and a noise burst, presented at varying stimulus onset asynchronies (Vatakis *et al.*, 2008). However, to examine the neural basis of cross-modal binding and to link it to perception requires the design of comparable behavioural paradigms, in which the percept influences the animal's behavioural response.

The task of grouping sensory inputs so as to generate a coherent percept corresponding to an individual source is not only a problem across sensory systems but also within them. In the latter case, the utility of animal models has been well established for studying auditory streaming and visual segmentation (Itatani & Klump, 2017). For example, ferrets previously have been shown to be capable of perceiving interleaved two tone sequences (ABAB, with a frequency shift of tone B) as separate streams and to detect regularly repeating target tones within a random

multitone background (Ma *et al.*, 2010). Similarly, it has been demonstrated that cats are able to use spatial cues to segregate two rhythmic patterns (Middlebrooks & Onsan, 2012) or detect a rhythm change between target and competing sound sequences (Javier *et al.*, 2016). In comparison, relatively few animal models have been developed for studying the more naturalistic scenario of cross-modal binding.

In this chapter, we describe a novel behavioural paradigm to evaluate cross-modal binding in the ferret. The ability of animals to categorise simultaneously presented auditory-visual stimuli as one object, in the spatially congruent stimulus condition, or as discrete objects, in the spatially incongruent condition, was tested by measuring their head orienting movements and approach-to-target responses.

5.2 Results

To investigate the ability of ferrets to categorise multisensory inputs, we devised a task that required them to make different behavioural responses depending on the spatial congruency of simultaneously presented auditory and visual stimuli (Fig. 2-5). On spatially congruent trials, auditory and visual stimuli were presented from the same location ([-30°A, -30°V] or [+30°V, +30°A]) and animals were rewarded for approaching the presentation site. In contrast, on spatially incongruent trials, auditory and visual stimuli were presented from separate locations ([-30°A, +30°V] or [-30°V, +30°A]) and animals were rewarded for approaching a neutral site (0°). Task performance was measured by recording the accuracy and timing of animals' initial head orienting movements and subsequent approach-to-target responses. Although, animals were only trained by operant conditioning to make an approach response, on most of the trials we also recorded a head orienting response, therefore both measures can be considered part of the categorisation behaviour. The short latency of the head orienting response and its unconditioned nature should provide an insight into the initial capture of attention by the different modalities at stimulus presentation, whereas the conditioned approach to target should provide an insight into the categorisation decision-making process. The results presented are based on the analysis of 7138 approach-to-target responses (see Table 2.1).

5.2.1 Task Performance

As described for the localisation task, auditory-visual stimulus presentation typically

elicited a head orienting response within 150 ms of stimulus onset. On most congruent trials, animals turned their heads in the direction of the stimulus location (-30° or +30°), which in this case corresponded to the intended response location, prior to leaving the central platform (Fig. 5-1, -30° and +30°). On some congruent trials, the amplitude of the head orienting movement matched the rewarded response location but, more commonly and particularly for the left side, it under-shot the target (mean final head bearing -30° vs. +30°: -14.7° vs. +21.4°). It is possible that this undershoot occurs in part due to eye movements contributing to larger gaze shifts that were more closely directed to the sound sources. Alternatively, provided that the amplitude of the head orienting response is sufficient to bring the sound source into the frontal region of the auditory system, where spatial acuity is sharpest, it is also possible that for this function the resolution of the head orienting response does not need to match that of the approach-to-target response. Initial inspection of the data from incongruent trials suggests that rather than consistently orienting towards the rewarded response location (0°), animals turned their heads towards the location of one of the unisensory stimulus components (-30° or +30°), as illustrated by Figure 5-1 (middle panel), which shows peaks on each side of the midline.

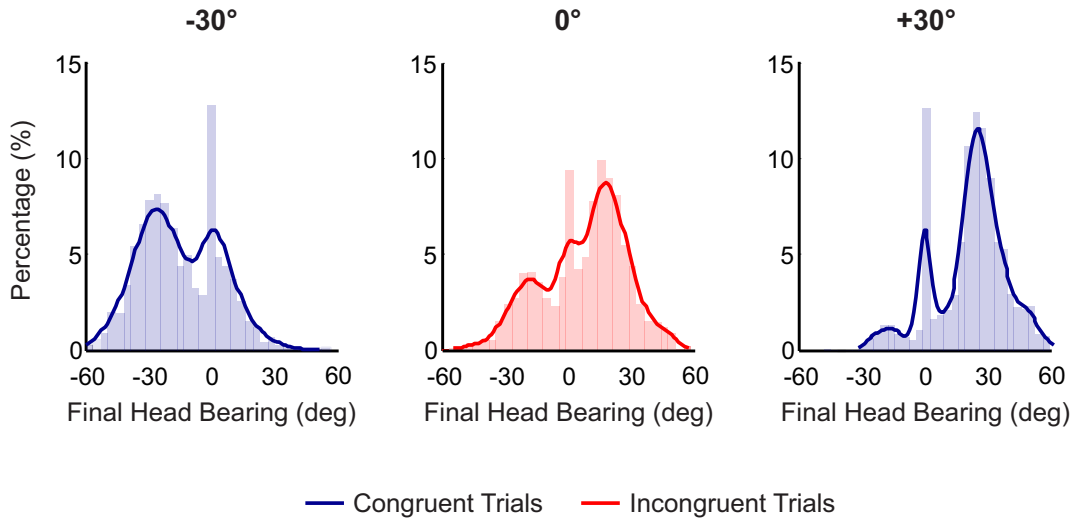


Figure 5-1. Distribution of final head bearings for spatially congruent (blue) and incongruent (red) stimuli. Histograms fitted with a kernel smoothing function (solid line) showing the distribution of final head bearings, binned in 3.75° intervals. Data are shown separately according to the rewarded response location (-30°; 0°; +30°).

However, closer inspection of the data from incongruent trials reveals a more complex relationship between the final head bearing and rewarded response location. In order to explore this relationship in more detail, individual animal head orienting data were analysed by plotting head bearing over time (Fig. 5-2A). This analysis illustrates that regardless of whether the auditory stimulus component was presented from either -30° or +30°, animals F1305, F1324 and F1404 oriented preferentially towards the right side, whereas animals F1440 and F1441 oriented preferentially towards the left side. However, it is also apparent from this figure that the amplitude of the head orienting movement towards the preferred side is greatest when the auditory stimulus component was also presented from this side.

In order to quantify this relationship, individual animal head orienting data were also analysed by plotting the mean final head bearing versus auditory stimulus location for congruent and incongruent trials (Fig. 5-2B). The slopes and y-intercepts

of the resulting lines provide a measure of the relative accuracy of the head orienting response, with perfect performance for congruent trials equating to a slope (gain) of 1, and a y-intercept (response bias) of 0, and for incongruent trials equating to a slope (gain) of 0, and a y-intercept (response bias) of 0 (see Table 5-1). In keeping with the response bias described above, the y-intercepts for animals F1305, F1324 and F1404 were positive, whereas for animals F1440 and F1441 they were negative, which is consistent with a right and left sided bias, respectively. Although a response bias exists, positive slope values for incongruent trials were observed across all animals, which implies that the head orienting movement was modulated more by the auditory than the visual stimulus component. Furthermore, this was the case even when the incorrect trials are omitted from the analysis (slope range: 0.03 – 0.12). Therefore, although the biphasic distribution of final head bearings arises principally because of an animal's individual preference to orient and then leave the central platform from a particular side, it also arises because their head orienting responses are more modulated by the auditory than the visual stimulus component. This is consistent with the superior auditory localisation performance described in Chapter 3.

Although animals made a head turn on the incongruent trials, final head bearing values were smaller for incongruent trials ([-30°A +30°V]: +1.30° ± 10.26; [+30°A -30°V]: +11.14° ± 15.63) than for congruent trials (taking the auditory component as reference) ([-30°A -30°V]: -14.07° ± 5.15; [+30°A +30°V]: +21.19° ± 14.06), indicating that the spatial incongruency of the visual stimulus component has an effect on the otherwise auditory-biased head oriented response. A linear mixed effects analysis fitted to the final head bearing data confirmed that this effect was significant across stimulus location-duration combinations (Fig. 5-2 C-D) ($P < 0.001$,

see appendix 5.5.1.1 for significance values) and, in addition, that final head bearings were not significantly affected by the stimulus duration (t_{5963} = -0.99, P = 0.32).

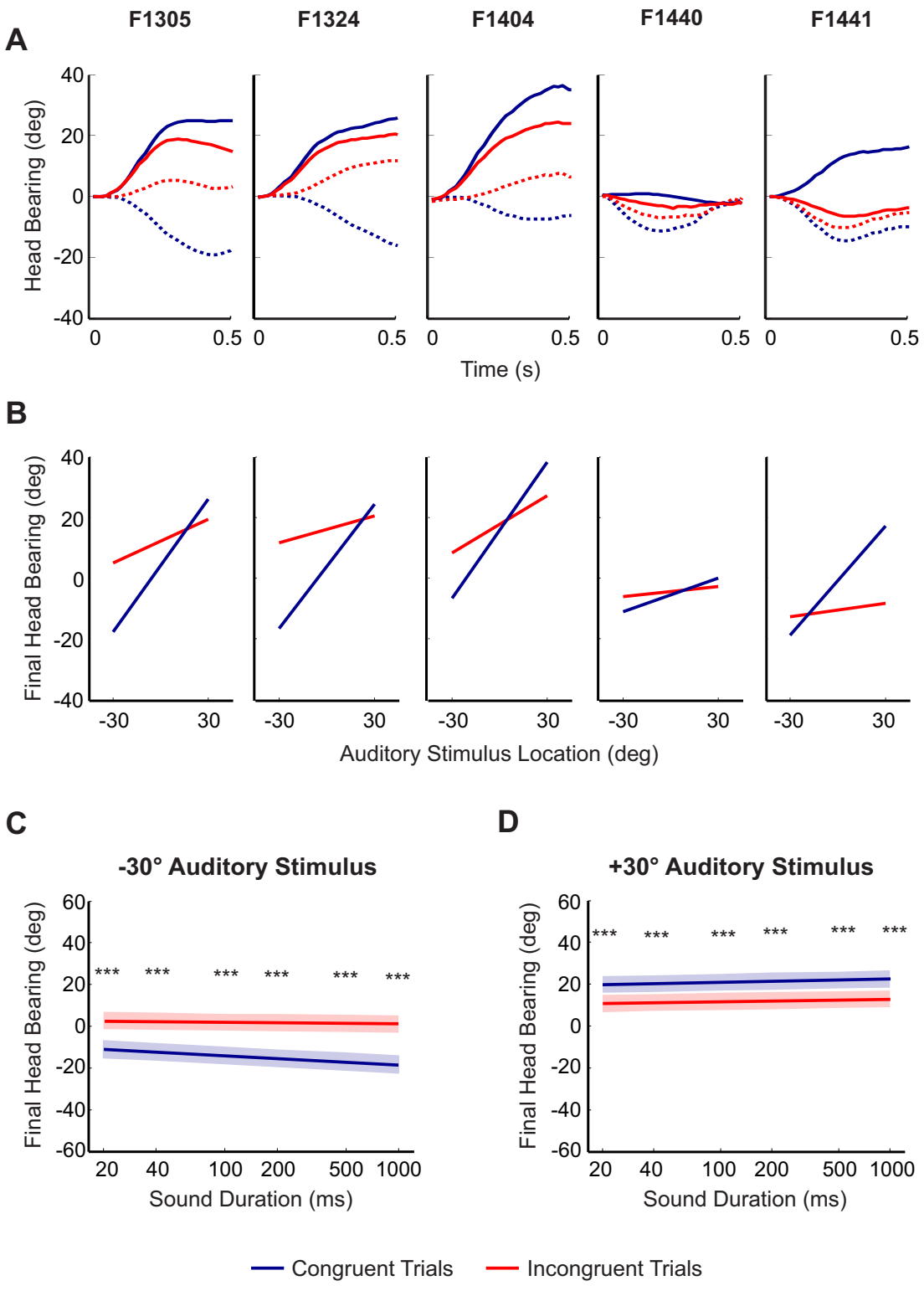


Figure 5-2. Comparison of final head bearings for spatially congruent (blue) and incongruent (red) stimuli. (A) Head bearings for the individual animals are shown as a function of time. Data are separated by trial congruency and by the location of the auditory stimulus component (-30°, dotted line; +30°, solid line). (B) Mean final head bearing plotted against the auditory stimulus component location for congruent and incongruent trials (see Table 5-1 for the slopes and y-intercepts of the resultant lines).

Data for individual animals are shown separately (C, D) Modelled final head bearings (± 95% CI) shown as a function of stimulus duration for congruent and incongruent stimuli. Significant differences between stimulus conditions are indicated by the asterisks and probability values by their number (* $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$). Data are shown separately according to whether the auditory component of the stimulus was presented from -30° (C) or +30° (D). Note that final head bearings were more modulated by congruent stimuli than by incongruent stimuli across stimulus duration - sound location combinations.

	Congruent Trials		Incongruent Trials	
	Slope	Y-Intercept	Slope	Y-Intercept
F1305	0.73	4.22	0.24	12.17
F1324	0.69	3.86	0.14	15.98
F1404	0.75	15.82	0.31	17.79
F1440	0.18	-5.41	0.05	-4.44
F1441	0.59	-0.69	0.08	-10.41

Table 5-1. Slopes and y-intercepts for the head orienting responses. The slopes and y-intercepts of lines fitted to the plots of final head bearing versus auditory stimulus location (Fig. 5-2B) are shown for each animal (rows) and stimulus type (columns).

Approach-to-target data were analysed by calculating probability correct scores and by fitting a linear regression model to confusion matrices quantifying the relationship between rewarded and actual response location (Fig. 5-3 A-C). For both congruent and incongruent stimuli, probability correct scores improved as stimulus duration increased, which was mirrored by an increase in the slope of the linear fits. Our results show that even at the shortest stimulus durations, animals were able to categorise stimuli at above chance levels. However, a clear difference was found in categorisation performance between the different stimulus conditions (Fig. 5-3 B-C), whereby for all stimulus durations the highest scores were obtained for congruent stimuli. To investigate the significance of this effect, a generalised linear mixed

effects model for Bernoulli-distributed responses was fitted to the approach-to-target data. The model shows that animals were significantly more likely to correctly categorise stimuli at longer durations ($t_{6892} = 5.4$, $P < 0.001$) and, in addition, congruent as compared to incongruent stimuli across all stimulus combinations ($P < 0.001$, see Fig. 5-3 B-C and appendix 5.5.1.2 for significance values).

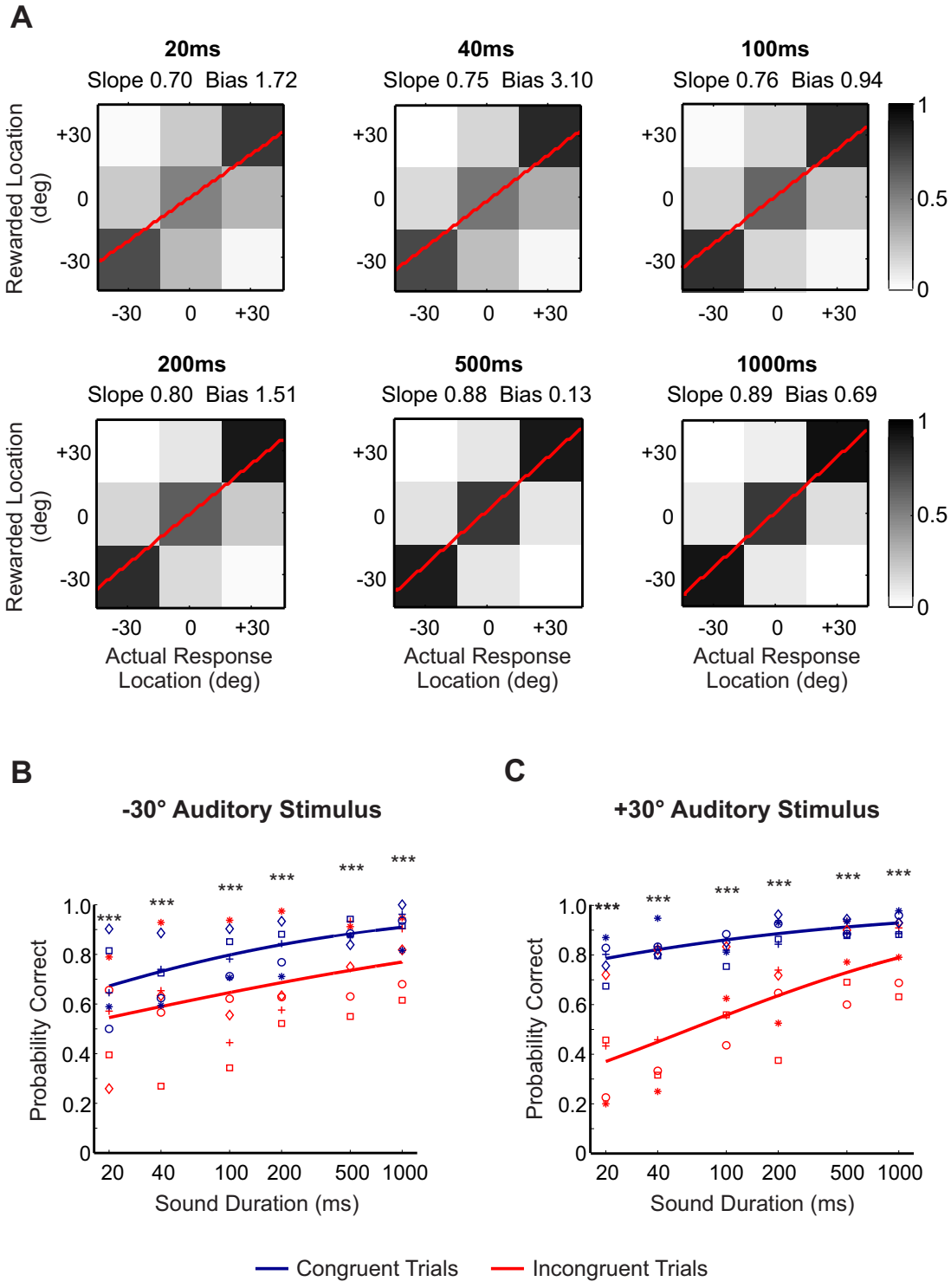


Figure 5-3. Comparison of approach-to-target performance for spatially congruent and incongruent auditory-visual stimuli. (A) Confusion matrices illustrating the conditional probability (represented in grey scale) of the actual response location for each of the rewarded response locations. Data for each stimulus duration (20 – 1000 ms) are shown in separate panels. A regression line was fitted (red line) to each confusion matrix, using a linear least squares method, and the corresponding slope (gain) and y-intercept (bias) are shown above each plot. (B-C) Modelled probability of a correct approach-to-

target response plotted as a function of stimulus duration (on a log scale) for spatially congruent (blue) and incongruent (red) stimuli. Data from the individual animals are represented by the different symbols. Significant differences between stimulus conditions are shown by the black asterisks and probability values by their number (* $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$). Note that approach-to-target categorisation performance was greater for congruent stimuli than for incongruent stimuli across all stimulus duration-location combinations (C, -30°A; D, +30°A).

Head orienting and approach-to-target data from incorrect incongruent trials were analysed further to investigate the interplay of attention capture between the modalities during task performance (Fig. 5-4A-C). The former was explored by calculating the final head bearing error relative to the location of each of the unisensory stimulus components, i.e. the difference between the final head bearing and the auditory and visual stimulus locations, respectively (Fig. 5-4A-B). The smallest errors were observed in relation to the auditory as opposed to the visual stimulus component (auditory = -18.4 ± 2.2°; visual = -41.6 ± 2.2°), indicating that the ferrets' head orienting movements were typically biased towards the auditory stimulus when presented simultaneously with the visual stimulus. Likewise, a higher percentage of incorrect responses on incongruent trials were made to the auditory stimulus location, which shows that this bias towards the sound source was also reflected in the subsequent approach-to-target response (auditory = 85.9% ± 4.7; visual = 14.1% ± 4.7) (Fig. 5-4C). Despite that the proportion of incorrect incongruent trials diminished at longer stimulus duration (Fig. 5-3), the bias towards the auditory stimulus remained and was even more pronounced (Fig. 5-4). Furthermore, the localisation task results show that approach-to-target performance for visual targets approached that for auditory targets at longer stimulus durations (20 ms: A-V = 25%; 1000 ms: A-V = 12%). Together, these results suggest that the

response bias observed towards the auditory stimulus component cannot be explained solely in terms of a difference in saliency between the stimulus modalities.

Although head orienting and approach-to-target behaviours appear to be highly correlated in terms of the direction of the initial head orienting movement and the location of the target subsequently approached, there are fundamental differences between them both in terms of the way they are affected by different stimulus parameters and the likely neural circuitry involved. The different parametric natures of each behaviour, continuous for head orienting and categorical for approach-to-target, can be used to highlight the discrepancies between them in the current task. For example, the distribution of the relative final head bearing errors for incorrect incongruent trials peaks at -30° (Fig. 5-4A), which indicates that on some trials animals oriented towards the rewarded location (0°), but still went on to approach an incorrect response location. Although this discrepancy could be explained in terms of both modalities failing to capture the animal's attention and, consequently, the animal making a random response, it can also be explained in terms of the orienting and approach-to-target response being independent of each other. In support of the second hypothesis, during correct incongruent trials animals typically oriented more towards the auditory stimulus component, which was presented from either -30° or + 30° (Fig. 5-2A-B), but still made a correct approach-to-target response at 0°. In summary, although our results suggest that the unconditioned head orienting movement might contribute to the categorisation process by providing additional dynamic cues, it seems likely that the categorisation decision occurs later in the response and, therefore, most likely depends on the cortical activity.

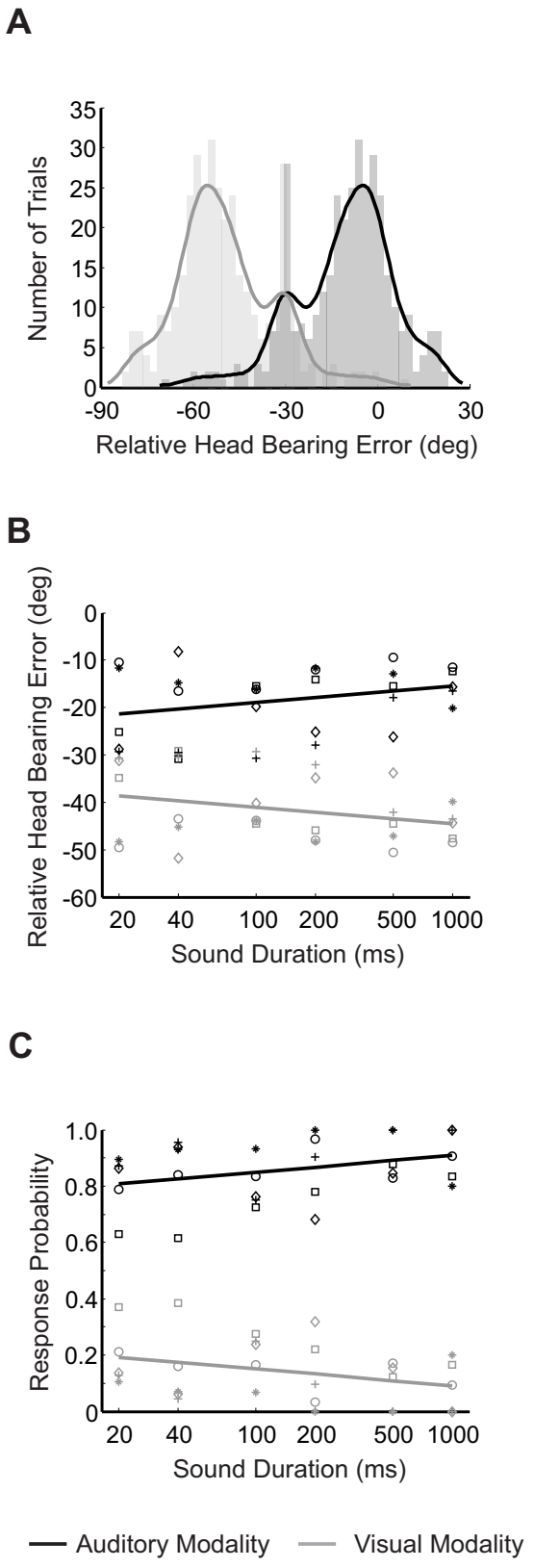


Figure 5-4. Head orienting and approach-to-target behaviour on incorrect incongruent trials. Only data from incongruent trials in which animals made an incorrect approach-to target response are shown. (A-B) Distribution of final head bearing errors binned in

2.5° intervals (A) and mean final head bearing errors shown as a function of stimulus duration (B, on a log scale). Final head bearing errors are shown relative to the location of either the auditory (black) or visual (grey) stimulus component. (C) Proportion of approach-to-target responses to either the auditory (black) or visual (grey) stimulus component, shown as a function of stimulus duration (on a log scale). Overall, animals typically orient towards and subsequently approach the auditory stimulus.

5.2.2 Reaction and Response Times

In previous reports of ferret operant behaviour, response times have been correlated with task difficulty (Nodal *et al.*, 2008; Bajo, Nodal, Moore, *et al.*, 2010; Leach *et al.*, 2013; Gold *et al.*, 2015). Therefore, head orienting reaction times and approach-to-target response times were analysed to assess whether ferrets also react and respond more rapidly to congruent than to incongruent stimuli. In view of the negative correlation between response times and the correctness of the approach-to-target response described for the localisation task (see Chapter 3.2.2), we first established whether the same effect applies to the categorisation task by fitting a generalised linear model to the approach-to-target response correctness data (Fig. 5-5A-B). The results show that head orienting reaction times ($t_{4736} = 5.20$, $P < 0.001$) and approach-to-target response times ($t_{6869} = -5.59$, $P < 0.001$) were significantly negatively correlated with the probability of a correct approach-to-target response regardless of the trial type (i.e. congruent or incongruent). In view of this result, together with the varying proportion of correct trials observed between stimulus conditions, incorrect trials were omitted from the analysis. In addition, to prevent any bias due to a possible lack of attention, which might have resulted in abnormally slow responses, reaction times > 250 ms and response times > 3.5 s were omitted

from the analysis. In total, 3.2 % of reaction times and 1.5 % of response time were omitted because of these criteria.

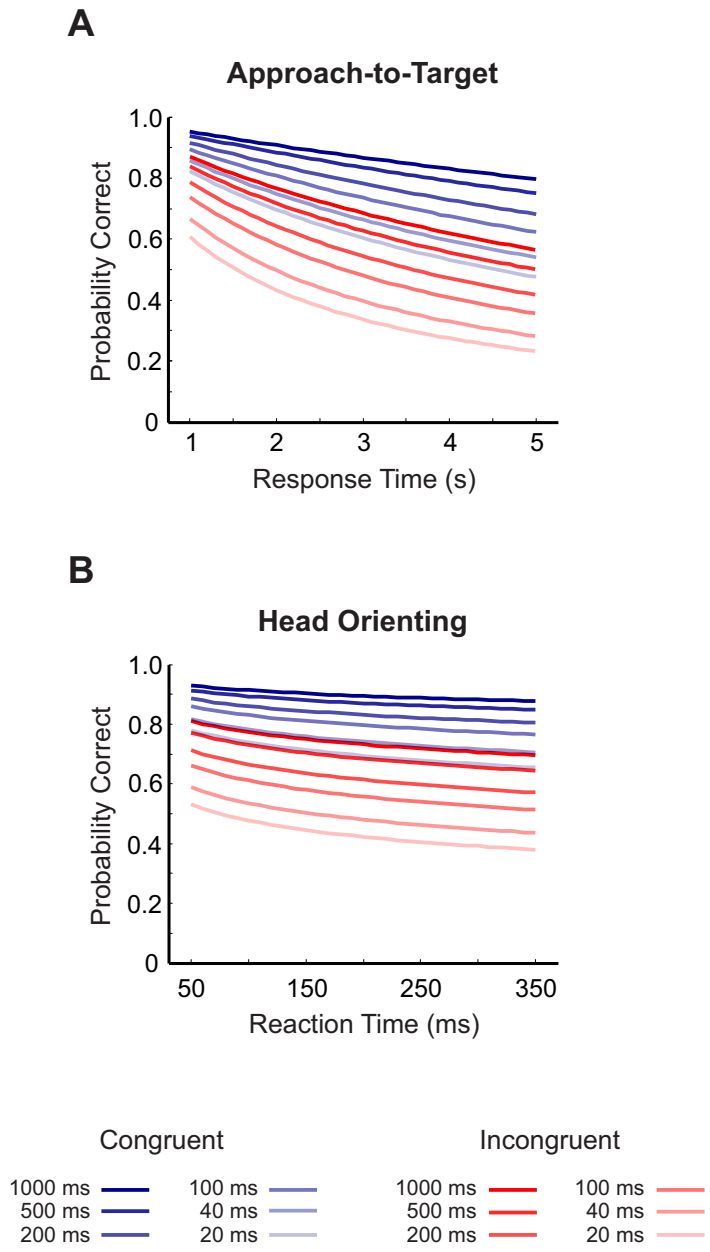


Figure 5-5. Approach-to-target categorisation accuracy versus reaction and response times. Modelled probability of a correct approach-to-target response plotted against approach-to-target response (A) and head orienting reaction (B) times for congruent (blue) and incongruent (red) stimuli. For both stimulus types, incorrect responses were associated with longer reaction and response times than correct responses.

In keeping with the unconditioned nature of the head orienting movement, reaction times for congruent and incongruent stimuli were similar (congruent vs. incongruent: 86.38 ± 14.40 vs. 90.79 ± 15.65 ms) (Fig. 5-6A). Furthermore, a linear mixed effects model confirmed that across the majority of stimulus combinations the spatial congruency of stimuli did not significantly affect reaction times (see Fig. 5-6B-C and appendix 5.5.1.4 for significance values). Moreover, the significant difference that was observed between trial types, when the auditory stimulus component was presented from +30° at the shortest stimulus durations, likely arises because animals F1440 and F1441 did not make a clear head turn on incongruent trials, when the auditory stimulus component was presented from the left side (Fig. 5-2A). For the localisation task, reaction times for auditory and auditory-visual stimuli were comparable and both were much shorter than those for visual stimuli (AV: 78.82 ± 0.55 ms, A: 87.05 ± 0.63 ms, V: 142.10 ± 1.01 ms), implying that the orienting movement was triggered by the auditory rather than the visual stimulus component. Likewise, reaction times from the categorisation task were similar to those for auditory-visual and auditory stimuli from the localisation task, suggesting that the same is true for head orienting movements during the categorisation task.

In contrast, approach-to-target response times were shorter for congruent stimuli (1.74 ± 0.23 s) than for incongruent stimuli (1.83 ± 0.25 s) (Fig. 5-7A). A linear mixed effects model fitted to the response time data confirmed that this effect was significant across all stimulus location-duration combinations (see Fig. 5-7B-C and appendix 5.5.1.3 for significance values). The model also shows that response times significantly increased at longer stimulus durations regardless of trial type (significance level range across stimulus combinations: congruent t_{5244} = -3.53 - -

1.99, $P < 0.001 – 0.05$; incongruent $t_{5244} = - 5.87 - -2.44$, $P < 0.001– 0.02$). Overall, these data suggest that the increase in categorisation accuracy with increasing stimulus duration is linked to the length of time available for information to be accumulated during the stimulus presentation, thereby resulting in a trade off between response time and accuracy, particularly for the incongruent trials. In fact, comparison of response times across trial types shows that they mirror the difficulty of the categorisation task, whereby incongruent stimuli took longer to categorise than congruent stimuli across animals and all stimulus location-duration combinations.

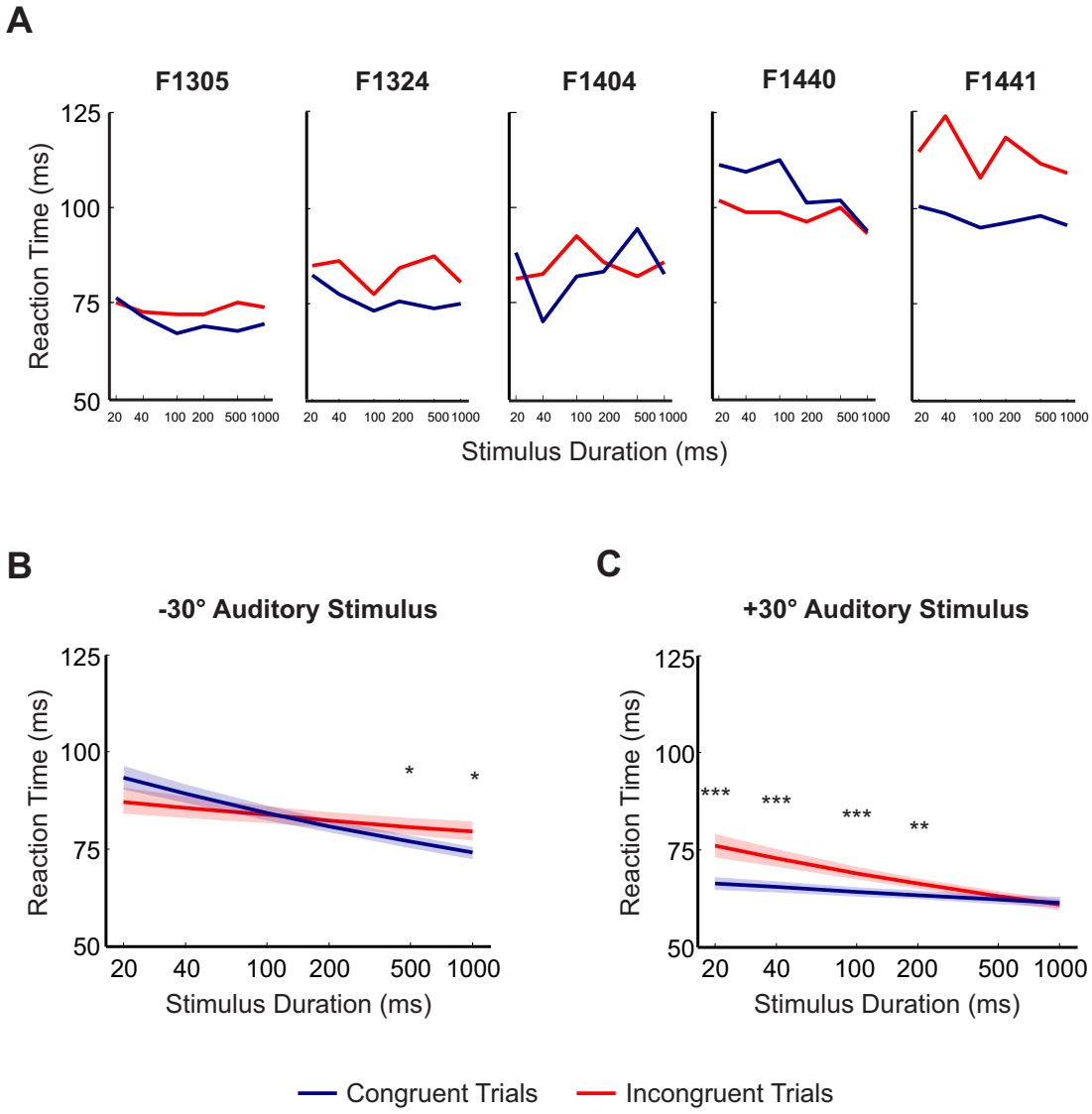


Figure 5-6. Comparison of head orienting reaction times for spatially congruent (blue) and incongruent (red) auditory-visual stimuli. (A) Reaction times shown as a function of stimulus duration (on a log scale) for congruent and incongruent trials. Individual animals are plotted on separate panels. (B-C) Modelled reaction times (± 95% CI) shown as a function of stimulus duration (on a log scale) for congruent and incongruent trials. Model results are shown on separate panels according to whether the auditory component of the stimulus was presented from -30° (B) or +30° (C). Significant differences between congruent and incongruent stimuli are indicated by the asterisks and probability values by their number (* $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$).

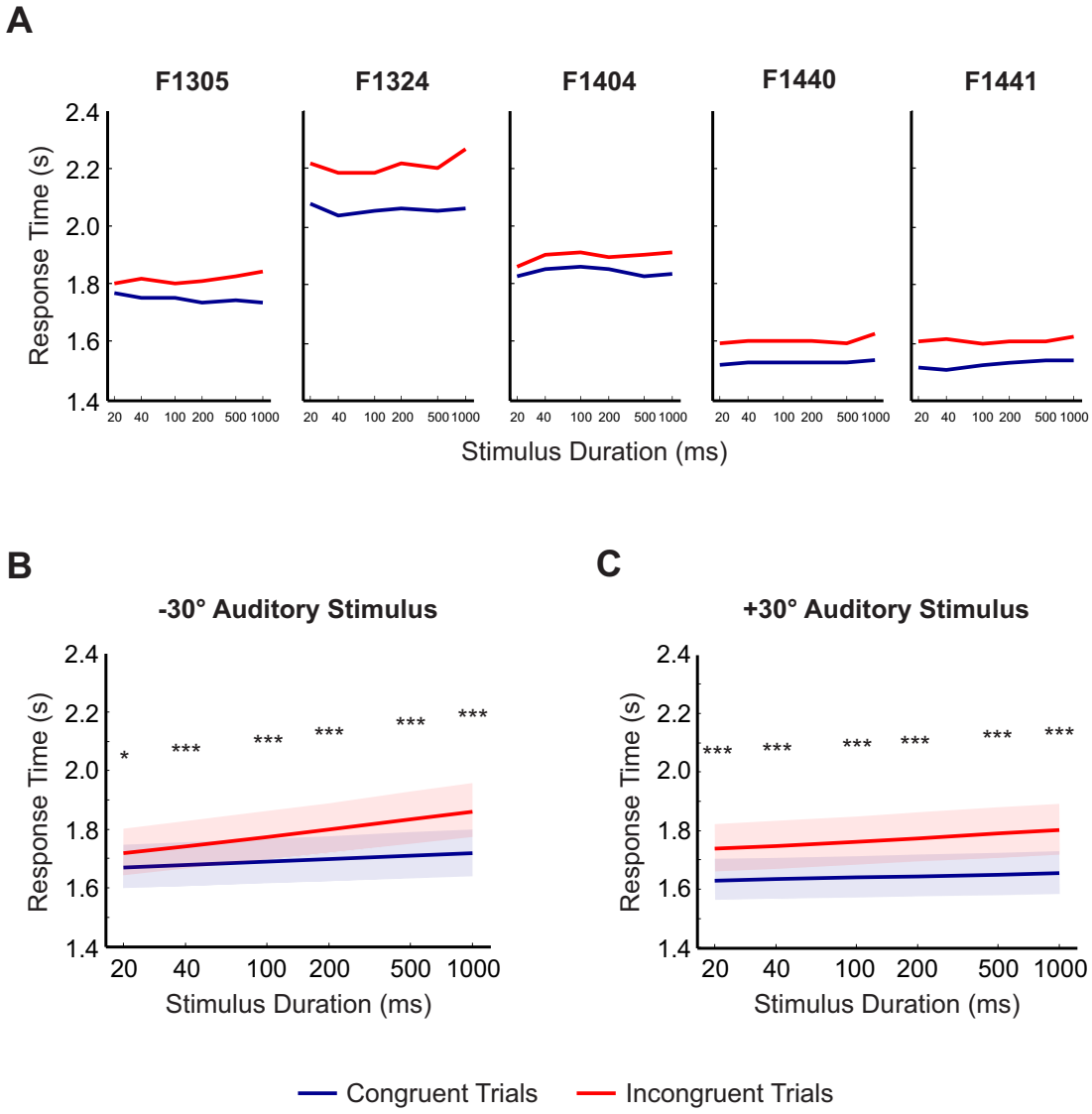


Figure 5-7. Comparison of approach-to-target response times for spatially congruent (blue) and incongruent (red) auditory-visual stimuli. (A) Response times shown as a function of stimulus duration (on a log scale) for congruent and incongruent stimuli. Individual animals are plotted on separate panels. (B-C) Modelled response times (± 95% CI) shown as a function of stimulus duration (on a log scale) for congruent and incongruent stimuli. The results are shown on separate panels according to whether the auditory component of the stimulus was presented from -30° (B) or +30° (C). Significant differences between congruent and incongruent stimuli are indicated by the asterisks and probability values by their number (*$P < 0.05$, ** $P < 0.01$, *** $P < 0.001$).

5.2.3 Comparison of Localisation and Categorisation Task Performance

Finally, to examine whether the context of behavioural task affects performance, we compared the accuracy and timing of the head orienting and approach-to-target responses made during the localisation and categorisation tasks. In terms of stimulus composition and reward location, ±30° auditory-visual trials from the localisation task were identical to spatially congruent trials from the categorisation task and, therefore, only data from ±30° auditory-visual trials were included in the analysis.

On the left side (-30°), final head bearing errors were significantly smaller (loc vs. cat: -12.03° ± 12.73 vs. -16.47° ± 5.92, t_{5969} = -2.64, P = 0.001) and approach-to-target probability correct scores were significantly higher (loc vs. cat: 0.83 ± 0.09 vs. 0.80 ± 0.08, t_{6956} = -2.35, P = 0.02) for the localisation task than for the categorisation task (Fig. 5-8 A-B). This would suggest that animals found the localisation task easier than the categorisation task, despite differences between the behavioural paradigms that might be expected to increase the difficulty of the localisation task. First, the scope of spatial attention differed between tasks since for the localisation task animals were required to attend to the frontal 180° field, whereas for the categorisation task they were only required to attend to the frontal 60° field. Second, the number of possible response locations was higher in the localisation task than in the categorisation task: 7 vs. 3, respectively. On the other hand, localisation appears to be a more natural behaviour than categorisation, as the respective times required to train the animals on each task indicates. In contrast, on the right side (+30°), approach-to-target probability correct scores (loc vs. cat: 0.88 ± 0.06 vs. 0.86 ± 0.04, t_{6956} = -0.97, P = 0.33) and final head bearing errors were not significantly

different (loc vs. cat: -10.19° ± 14.07 vs. -9.03° ± 13.98, $t_{5969} = 0.01$, $P = 0.92$).

On the other hand, the head orienting reaction and approach-to-target response times were significantly shorter on the right side (+30°) for the categorisation task than for the localisation task (Fig. 5-8C-D) (loc vs. cat reaction times: 82.43 ms ± 27.72 vs. 76.20 ms ± 13.60, $t_{4494} = 6.32$, $P < 0.001$; loc vs. cat response times: 1.70 s ± 0.12 vs. 1.68 s ± 0.11, $t_{5841} = 2.50$, $P = 0.01$), whereas on the left side (-30°) no significant difference was identified between tasks (loc vs. cat reaction times: 96.15 ms ± 37.53 vs. 105.87 ms ± 25.96, $t_{4494} = -1.44$, $P = 0.15$; loc vs. cat response times: 1.77 s ± 0.33 vs. 1.75 s ± 0.25, $t_{5841} = 0.13$, $P = 0.89$). This difference could reflect that the congruent trials from the categorisation task resemble a lateralisation rather than a localisation task, therefore a simple left or right stereotyped head orienting response without further amplitude modulation would suffice. Whereas, in the localisation task further amplitude modulation would be advantageous to discriminate between different sources within the same hemifield.

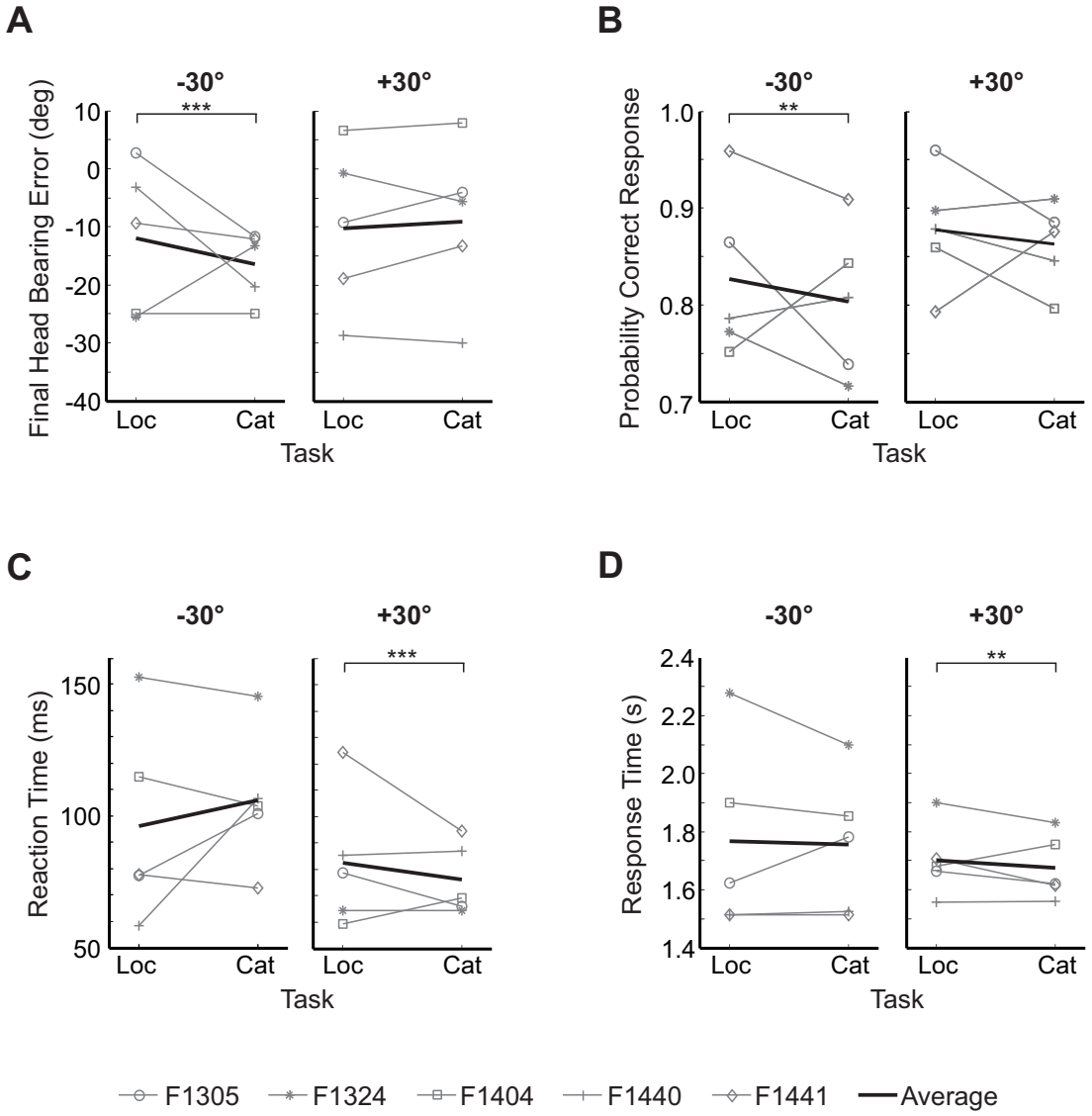


Figure 5-8. Comparison of localisation and categorisation task performance. (A) Final head bearing errors, (B) probability of a correct approach-to-target response, (C) reaction times and (D) response times are shown for the localisation (Loc) and categorisation (Cat) tasks. For both tasks, only data from spatially congruent auditory-visual trials with stimuli presented from either -30° or +30° are included. Significant differences between the localisation and the categorisation tasks are indicated by the asterisks and probability values by their number (*$P < 0.05$, ** $P < 0.01$, *** $P < 0.001$).

5.3 Discussion

The task of grouping sensory inputs so as to generate a coherent percept corresponding to an individual source is not only a challenge within each sensory system but also across them. Whilst a variety of behavioural studies have evaluated various aspects of either auditory scene analysis or visual image segmentation using animal models (Itatani & Klump, 2017), relatively few have focused on the more naturalistic scenario, when auditory-visual stimuli are combined. Therefore, in the current chapter, we have described a novel behavioural paradigm to evaluate cross-modal binding in the ferret, whereby the ability of ferrets to categorise simultaneously presented auditory-visual stimuli as one object, in the spatially congruent stimulus condition, or as discrete objects, in the spatially incongruent stimulus condition, was characterised by measuring the accuracy and timing of their initial head orienting movements and subsequent approach-to-target responses.

Ferrets previously have been shown to be able to perceive two tone sequences (ABAB sequence, with a frequency shift of tone B) as separate streams and to detect regularly repeating target tones within a random multi-tone background (Ma *et al.*, 2010). In other species, spatial separation has also been shown to provide cues for stream segregation (O'Connor & Sutter, 2000; Boehnke & Phillips, 2005; Eramudugolla *et al.*, 2008). For example, human listeners can segregate two rhythmic patterns if the sources differ spatially by as little as 8° (Middlebrooks & Onsan, 2012). Likewise, cats were found to be capable of detecting a rhythm change when the target and competing sequence sound bursts were separated by 9.4° (Javier *et al.*, 2016). The results presented in the present chapter show that ferrets can also

use spatial cues for stream segregation across sensory systems, since they were able to categorise, as indicated by their approach-to-target responses, light flashes and noise bursts separated by 60° in the horizontal plane as discrete objects.

Nevertheless, we found that approach-to-target categorisation responses were more accurate and faster to spatially congruent stimuli than to incongruent stimuli. This discrepancy in performance between trial types likely arises due to a combination of factors pertaining to the training of animals, stimulus and task design. First, all animals were initially trained to localise auditory-visual stimuli and subsequently tested on the localisation task prior to the categorisation task. Second, to ensure equal probability of a reward at each spout, the ratio of congruent to incongruent trials was 2:1. Third, it seems likely that the perceptual load of the categorisation decision differed between the trial types given that on incongruent trials animals had to resolve a conflict between stimulus locations. Finally, the difficulty of indicating the categorisation decision also varied between trial types. Specifically, on congruent trials, animals indicated their decision by approaching the stimulus location, which is in keeping with their natural phonotaxic behaviour and prior training. However, on incongruent trials animals indicated their decision by approaching a neutral site, which required an added conflict between the stimulus and reward location to be resolved after the categorisation decision was made.

The analysis of error data from incorrect incongruent trials provides an insight into the interplay of attention capture between the modalities during task performance and showed that animals overwhelmingly oriented towards and, in addition, approached the auditory rather than the visual stimulus component. This is consistent with our localisation task finding that the latency and accuracy of animals

head turns to spatially congruent auditory-visual stimuli were not significantly different from those made to sound alone. In contrast to our results, visual rather than auditory dominance has been described extensively in other species (Miller, 1973; Colavita, 1974; Randich *et al.*, 1978; Sinnett *et al.*, 2007). For example, when humans are presented with an auditory, visual or auditory-visual target and instructed to press one of three corresponding response keys, the majority of errors on bimodal trials were comprised of erroneously pressing the visual rather than the auditory response key (Sinnett *et al.*, 2007). Furthermore, subsequent studies have shown that this was also the case, albeit to a lesser extent, even when the unisensory stimulus components are presented from separate locations (Koppen & Spence, 2007).

In the present study, it is unlikely that the auditory dominance observed was a result of the way animals were trained, since all animals were trained initially using auditory-visual stimuli. It is well established that an aspect of a scene can capture attention because of either bottom-up processing (i.e. its strong physical features) or top-down processing (i.e. it is a goal directed priority for successful behaviour) (Kondo *et al.*, 2017). Therefore, at shorter stimulus durations, the auditory dominance observed during the categorisation task might simply arise because the noise burst provided a more salient cue than the LED. However, our results suggest that this effect was offset at longer stimulus durations. For example, our results from the localisation task indicate that at longer stimulus durations approach-to-target performance for visual targets approached that for auditory targets. It is possible to explain the auditory dominance observed at longer stimulus durations in terms of optimal cue integration. For example, human psychophysics studies, have shown that the ventriloquism effect can be explained in terms of optimal cue integration,

whereby cues are weighted according to their reliability and thus vision dominates the percept because it specifies the stimulus location more precisely than audition (Alais & Burr, 2004). Furthermore, if the visual stimuli is blurred then the this effect has been shown to reverse so that audition rather than vision dominates (Alais & Burr, 2004). Likewise, our results from Chapter 4 suggest that ferrets also integrate information from different sensory modalities in a statistically optimal manner. Therefore, it is possible that the auditory dominance observed in the categorisation task arises because audition provides more reliable information than vision and therefore it is given greater weight in the categorisation decision. From an ethological perspective this would seem intuitive, given that ferrets are adapted for hunting in dark burrows and would therefore naturally rely on audition over vision.

An interesting question to consider is at what point the categorisation decision occurs during the trial. Depending on the distance of the source from the subject, at the early stages of multisensory processing a significant temporal disparity can exist between auditory and visual inputs. This disparity arises because although propagation is faster for light than for sound, this is outweighed by transduction being much slower for visual stimuli than for auditory stimuli. For example in the cat SC, mean response latencies for auditory and visual stimuli presented at a distance of 15 cm have been shown to be 14 ms and 91 ms, respectively (Wallace & Stein, 1996). Theoretically, for the aforementioned physical and physiological effects on response latency to cancel out and, thus, for temporal coincidence between inputs to occur, stimuli would need to be presented at a distance of ~ 20 m (Lewald *et al.*, 2001). Therefore, it is perhaps not surprising that reaction times for congruent and incongruent stimuli are similar to those for auditory stimuli in the localisation task,

suggesting that stimulus detection was driven primarily by the auditory modality. Furthermore, this initial detection bias was also reflected in the subsequent head orienting movement, given that on incongruent trials animals' head turns were typically modulated more towards the auditory than the visual stimulus component.

In addition to focusing attention, the orienting response has been postulated to play an active part in decision-making. For example, recordings from the SC of the rat during an odour guided spatial choice task suggest that neural activity at this level can even represent the decision variable, given that the majority of neurons encoded choice direction during and even after locomotion (Felsen & Mainen, 2008). Our results show that although animals tend to look in the direction of the auditory stimulus component, the amplitude of the head movement was smaller for incongruent than for congruent trials, suggesting that head orienting might also contribute to the categorisation decision-making process. This difference in the amplitude of the head orienting movement between trial types may result from the summation of signals within the SC given that, on incongruent trials, auditory and visual stimulus components were presented from different hemifields and would therefore each have stimulated the contralateral SC, respectively (Sahibzada *et al.*, 1986). However, although correlated, we also found that the head orienting response did not always determine the approach-to-target response, thus indicating an independence between these behaviours. In summary, although our results suggest that the unconditioned head orienting movement might contribute to the categorisation process by providing additional dynamic cues, it seems likely that the decision occurs later in the response and therefore depends on the cortical activity.

The categorisation task results show that ferrets were able use spatial cues to

merge different sensory stimuli and resolve conflicts between them. Likewise, at the single neuron level, spatial cues also have been shown to be a key determinant of multisensory integration in the SC (Meredith & Stein, 1986a) and cortical association areas (Wallace *et al.*, 1992). However, the neural basis for auditory-visual stream segregation by spatial cues in early auditory cortex is not immediately intuitive given that in the visual system stimulus location is encoded by the distribution of activity across the receptor surface of the retina, whereas in the auditory system stimulus location is inferred from cues generated by the interaction of sound waves with the head and external ears (King et al, 2001). Simultaneous recordings while animals execute the categorisation task described in this chapter should provide insight into this mechanism and, in addition, the extent to which early cross-modal activity is tied to the physical characteristics of stimuli as compared to the behavioural context of the task or the animal's perception, as indicated by its behavioural choice.

5.4 Appendix

5.4.1 Statistical Analyses

5.4.1.1 Final Head Bearings

To investigate the effect of stimulus congruency on final head bearings a linear mixed effects model was fitted to the data. A mixed effects model was used to allow for within ferret correlation by introducing a random effect across ferrets. The R code for a single stimulus duration (20 ms) and location (-30°) combination reads:

```
FHB ~ Congruency*Sound_Loc*I(log(Duration) - log(20)), random = ~
1|Ferret, data = head)
```

Consecutive re-levelling of the model for stimulus location and adjustment of the intercept for stimulus duration yielded the results shown in Table 5-2.

Duration	**FHB_{Incon}**	**FHB_{Con}**	***t*-value**	***P*-value**
Sound location -30°				
20	2.60	-11.04	-11.30	<0.001
40	2.32	-12.32	-15.57	<0.001
100	1.95	-14.02	-22.97	<0.001
200	1.67	-15.31	-25.35	<0.001
500	1.30	-17.00	-20.99	<0.001
1000	1.02	-18.29	-17.16	<0.001
Sound Location +30°				

20	10.85	19.82	7.23	<0.001
40	11.23	20.30	9.42	<0.001
100	11.72	20.94	13.01	<0.001
200	12.09	21.42	13.87	<0.001
500	12.59	22.05	10.97	<0.001
1000	12.96	22.53	8.60	<0.001

Table 5-2. Results of the linear mixed effects fitted to final head bearing data. Modelled final head bearing (deg) for congruent (FHB_{Con}) and incongruent (FHB_{Incon}) stimulus conditions are shown for each stimulus location-duration combination, together with the *P*-values for the following comparison: congruent vs. incongruent (con vs. incon).

A Shapiro-Wilk test was performed (W = 0.92776, $P = 0.5811$) and Q-Q plots (Fig. 5-9) were generated to check the goodness of the model fit, which both showed that the residuals were normally distributed and, thus, demonstrated a good fit.

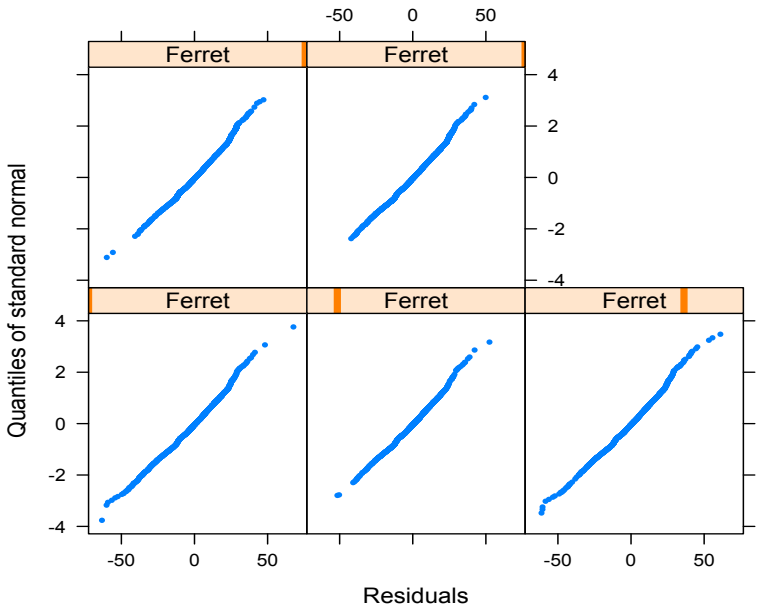


Figure 5-9. Q-Q plot of residuals from the linear mixed effects model fitted to the final head bearing data. Plots for individual animals are shown in separate panels. The points falling on a straight line indicate that the residuals are normally distributed.

5.4.1.2 Approach-to-Target Performance

To investigate the effect of stimulus modality on approach-to-target response correctness within stimulus durations and locations, a generalised linear model for Bernoulli distributed responses was fitted to the data. The R code and output for a single stimulus duration (20 ms) and location (-30°) combination reads:

```
mod.con <- glmmPQL(Correctness~Congruency*Sound_Loc*I(log(Duration)
- log(20)), random  = ~ 1|Ferret, family = binomial, data = Catt)
```

```
                                Value Std.Error   DF t-value p-value
(Intercept)                     0.182     0.162 6892   1.121   0.262
Congruency1                     0.537     0.138 6892   3.886   0.000
Sound_Loc30                    -0.711     0.161 6892  -4.414   0.000
I(log(Dur)-log(20))             0.261     0.049 6892   5.362   0.000
Cong1:Sound_Loc30               1.292     0.204 6892   6.340   0.000
Cong1:I(log(Dur)-log(20))       0.144     0.064 6892   2.260   0.024
Sound_Loc30:I(log(Dur)-log(20))
                                      0.211     0.071 6892   2.995
0.003
Cong1:Sound_Loc30:I(log(Dur) - log(20))
                               -0.291     0.094 6892  -3.104   0.002
```

The model output shows, for example, that the log(odds) corresponding to the congruent stimulus condition at -30° and 20 ms is 0.182 + 0.537. From this value, the probability (P_{Con}) of a correct response can be derived as follows:

$$\log\left(\frac{P_{Con}}{1 - P_{Con}}\right) = 0.182 + 0.537 = 0.719$$

$$\frac{P_{Con}}{1 - P_{Con}} = \exp(0.719) = 2.052$$

$$P_{Con} = 2.052(1 - P_{Con})$$

$$P_{Con} + 2.052P_{Con} = 2.052$$

$$P_{Con} = 0.672$$

Consecutive re-levelling of the model for stimulus location and adjustment of the intercept for stimulus duration yielded the results shown in Table 5-3.

Duration	P_{Con}	P_{Incon}	*t*-value	*P*-value
Sound Location -30°				
20	0.672	0.545	3.886	<0.001
40	0.731	0.590	5.962	<0.001
100	0.798	0.646	9.039	<0.001
200	0.839	0.686	9.362	<0.001
500	0.883	0.736	7.810	<0.001
1000	0.909	0.769	6.715	<0.001
Sound Location +30°				
20	0.785	0.370	12.216	<0.001
40	0.821	0.450	14.943	<0.001
100	0.861	0.558	17.440	<0.001
200	0.886	0.636	15.060	<0.001
500	0.913	0.729	9.924	<0.001
1000	0.929	0.789	7.116	<0.001

Table 5-3. Results of the generalised linear mixed effects model fitted to approach-to-target response correctness data. Modelled probabilities of a correct approach-to-target response for congruent (P_{Con}) and incongruent (P_{Incon}) stimuli are shown for each stimulus location-duration combination, together with the *P*-values and *t*-values (in brackets) for the following comparison: congruent vs. incongruent.

To check the fit of the generalised linear model, a half normal plot was generated (Fig. 5-10). Ordered absolute values of the Pearson residuals were plotted against the expected values for a binomial distribution and, in addition, a simulated envelope was generated, which corresponds to a 95% confidence interval. The model fit is good if the plotted points fall within the boundaries of the simulated envelope.

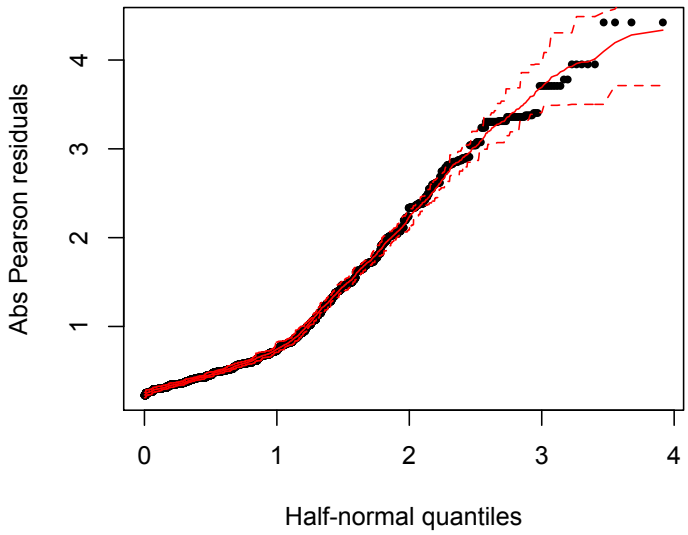


Figure 5-10. Half-normal plot of the Pearson residuals from the generalised linear model fitted to the approach-to-target data from the categorisation task. A simulated envelope (dashed lines) is also plotted together with the means of the simulated values (solid line). The points fall within the simulated envelope, indicating a good model fit.

5.4.1.3 Approach-to-Target Response Times

To investigate the effect of stimulus condition (i.e. congruent vs. incongruent) on approach-to-target response times within stimulus durations and locations, a linear mixed effects model was fitted to the data. Prior to fitting the model, response times were first transformed using a reciprocal function. The R code for a single stimulus

duration (20 ms) and sound location (-30°) combination reads:

```
mod.rt <- lme(1/Response_Time ~
Congruency*Sound_Loc*I(log(Duration)-log(20)), random = ~ 1|Ferret,
data = rts)
```

Consecutive re-levelling of the model for stimulus location and adjustment of the intercept for stimulus duration yielded the results shown in Table 5-4.

Duration	RT_{Incon}	RT_{Con}	*t*-value	*P*-value
Sound location -30°				
20	1.719	1.670	2.995	0.003
40	1.743	1.678	4.927	<0.001
100	1.775	1.690	8.690	<0.001
200	1.780	1.697	10.851	<0.001
500	1.834	1.709	10.205	<0.001
1000	1.861	1.717	8.990	<0.001
Sound Location +30°				
20	1.737	1.630	5.911	<0.001
40	1.748	1.678	7.955	<0.001
100	1.763	1.689	11.867	<0.001
200	1.774	1.697	13.995	<0.001
500	1.789	1.709	12.116	<0.001
1000	1.801	1.717	9.746	<0.001

Table 5-4. Results of the linear mixed effects model fitted to approach-to-target response times from the categorisation task. Response times (s) to congruent (RT_{Con}) and incongruent (RT_{Incon}) stimuli are shown for each stimulus duration and location, together with the *t* and *P*-values for the comparison: congruent vs. incongruent.

A Shapiro-Wilk test was performed (W = 0.9189, $P = 0.5229$) and Q-Q plots (Fig.

5-11) were generated to check the goodness of fit of the model, which both showed that the residuals were normally distributed and, thus, demonstrated a good fit.

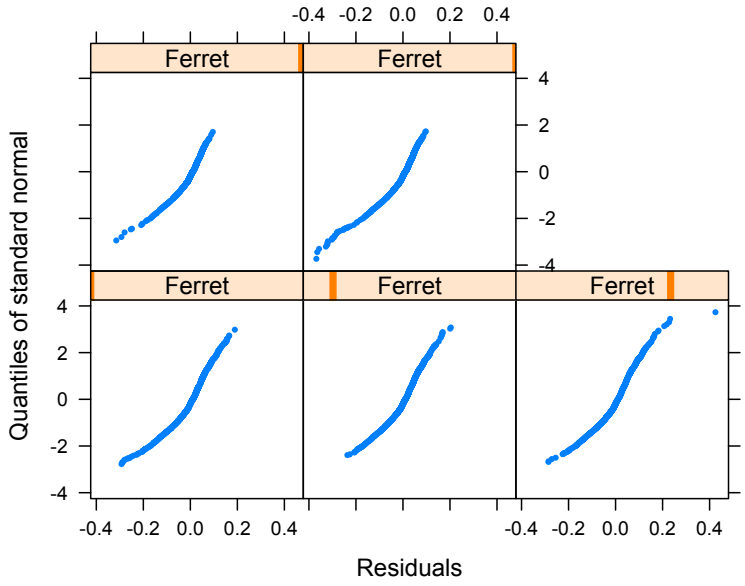


Figure 5-11. Q-Q plot of residuals from the linear mixed effects model fitted to the response time data. Plots for individual animals are shown in separate panels. The points falling on a straight line indicate that the residuals are normally distributed.

5.4.1.4 Head Orienting Reaction Times

To investigate the effect of stimulus congruency on head orienting reaction times within stimulus duration and location, a linear mixed effects model was fitted to the data. Prior to fitting, times were transformed using a reciprocal function. The R code for a single stimulus duration (20 ms) and sound location (-30°) combination reads:

```
mod.rt <- lme(1/Reaction_Time~Congruency*Sound_Loc*I(log(Duration)-
log(20)), random = ~ 1|Ferret, data = con.rt)
```

Consecutive re-levelling of the model for stimulus location and adjustment of the intercept for stimulus duration yielded the results shown in Table 5-5.

Duration	RT_{Incon}	RT_{Con}	*t*-value	*P*-value
Sound location -30°				
20	87.962	93.240	-1.600	0.111
40	85.609	89.135	-1.239	0.215
100	83.759	84.239	-0.254	0.799
200	82.406	80.873	0.956	0.339
500	80.691	76.823	2.103	0.036
1000	79.441	74.014	2.400	0.016
Sound Location +30°				
20	75.976	66.309	3.634	<0.001
40	72.759	65.385	3.737	<0.001
100	68.904	64.201	3.634	<0.001
200	66.252	63.335	2.808	0.005
500	63.040	62.224	0.720	0.472
1000	60.813	61.410	0.427	0.670

Table 5-5. Results of the linear mixed effects model fitted to head orienting reaction time data from the categorisation task. Reaction times (ms) to congruent (RT_{Con}) and incongruent (RT_{Incong}) stimuli are shown for each stimulus duration and location, together with the *t*- and *P*-values for the comparison: congruent vs. incongruent.

A Shapiro-Wilk test was performed (W = 0.95179, $P = 0.7499$) and Q-Q plots (Fig. 5-12) were generated to check the goodness of the model fit, which both showed that the residuals were normally distributed and, thus, demonstrated a good fit.

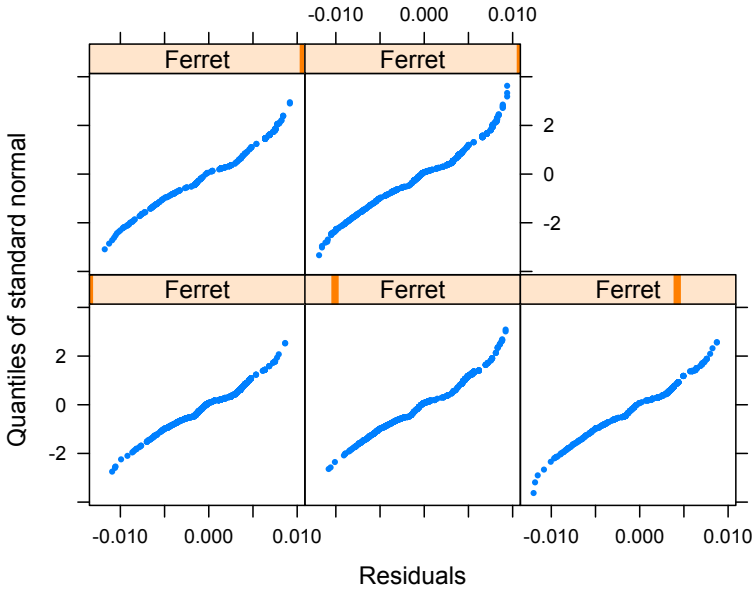


Figure 5-12. Q-Q plot of residuals from the linear mixed effects model fitted to the reaction time data. Plots for individual animals are shown in separate panels. The points falling on a straight line indicate that the residuals are normally distributed.

6 Neural Activity in the Auditory Cortex during Multisensory Categorisation

6.1 Introduction

The results from the different multisensory tasks described in this thesis have shown that the presentation of a stimulus from one modality can influence the perception or behavioural response to a stimulus from a different modality. Sensory illusions, including the ventriloquism effect, illustrate that the binding of multisensory stimuli is a fast and pre-attentive process (Thomas, 1941; Bushara *et al.*, 2003). However, the neural mechanisms that underlie multisensory integration in early sensory cortices, including the use of spatial cues to merge stimuli and resolve conflicts between them, remain unclear.

Multisensory stimuli that are produced by the same object tend to be approximately spatially and temporally coincident. Therefore, it is perhaps not surprising that whether or not neurons in the SC integrate inputs from different sensory modalities has been shown to depend on the physical properties of stimuli (Wallace *et al.*, 1992), whereby the greatest interactions occur when stimuli are proximate in time (Meredith & Stein, 1983; King & Palmer, 1985; Meredith *et al.*, 1987) and space (King & Palmer, 1985; Meredith & Stein, 1986a). Similar principles have also been reported for higher-order cortical areas, including the anterior ectosylvian sulcus (Wallace *et al.*, 1992) and ventral intra-parietal area (Schlack *et al.*, 2005). It is, however, not straightforward to extrapolate these principles to the auditory cortex, given that these regions fundamentally differ in their organisation.

In the SC, different sensory representations are topographically organised and aligned to form maps of space. Consequently, the different spatial receptive fields of neurons are in spatial register. Therefore, multisensory stimuli arising from the same location in space can modulate the activity of the same neuron. In the cortex, however, there is no such map of auditory space, and it is likely that sound source azimuth is encoded by a comparison of activity between neurons with heterogeneous spatial sensitivity (Day & Delgutte, 2013; Goodman *et al.*, 2013; Keating *et al.*, 2015). Therefore, it is unclear how retinotopic visual inputs, whether they originate subcortically or from other parts of the cortex, might "map" onto this arrangement.

In the present study, neural activity from early processing areas of the ferret auditory cortex has been recorded during the performance of a categorisation task, to investigate how the relative location of the stimuli affect their integration and categorisation at the level of the auditory cortex. The results presented in this chapter

also provide further insight into whether neural activity in the early auditory cortical processing areas is more closely tied to the physical properties of the stimulus, the context of the task or the animal's perception, as indicated by its behavioural choice.

6.2 Results

In this chapter, the results of the analysis of neural activity recorded from multi-electrode arrays implanted bilaterally over the auditory cortex during the performance of a modified version of the categorisation task are presented. To ensure that animals did not reach near perfect performance, stimulus duration was fixed at 100 ms. In contrast, sound level was roved pseudo-randomly across trials in 7 dB steps from either 69 – 90 dB SPL (F1305) or 20 – 55 dB SPL (F1440 and F1441), whilst the intensity of the visual stimulus was kept constant. The present results are based on the analysis of the neural responses recorded from three animals over 7788 trials.

6.2.1 Categorisation Task Performance

Behavioural data obtained during the neural recording session were analysed by constructing confusion matrices illustrating the relationship between the actual and the rewarded approach-to-target response location (Fig. 6-1A) and, in addition, by calculating probability correct scores (Fig. 6-1B-C). This analysis showed that, following implantation of electrode arrays, animals were still able to categorise congruent and incongruent stimuli at above chance levels. As for the pre-implantation behavioural results, a clear difference in performance was evident between stimulus types, whereby the highest probability correct scores were observed for congruent stimuli. A generalised linear model for Bernoulli-distributed responses

fitted to the approach-to-target data showed that this effect was significant for the majority of stimulus intensity-location combinations (see appendix 6.4.1.1 for significance values). As might be expected, given that experiments were conducted using a higher range of sound intensities (69 – 90 dB SPL), the highest percentage correct scores were observed for animal F1305, which was mirrored by an increase in the slope of the linear fit. For the neural recordings, the categorisation task was modified to include a control condition, whereby for one in every three incongruent trials a visual-visual ([-30°V +30°V]) stimulus was presented. The results show that animals were also able to correctly categorise the visual-visual stimulus at above chance levels (i.e. > 33%, F1305: 83%; F1440: 54% F1441: 42%).

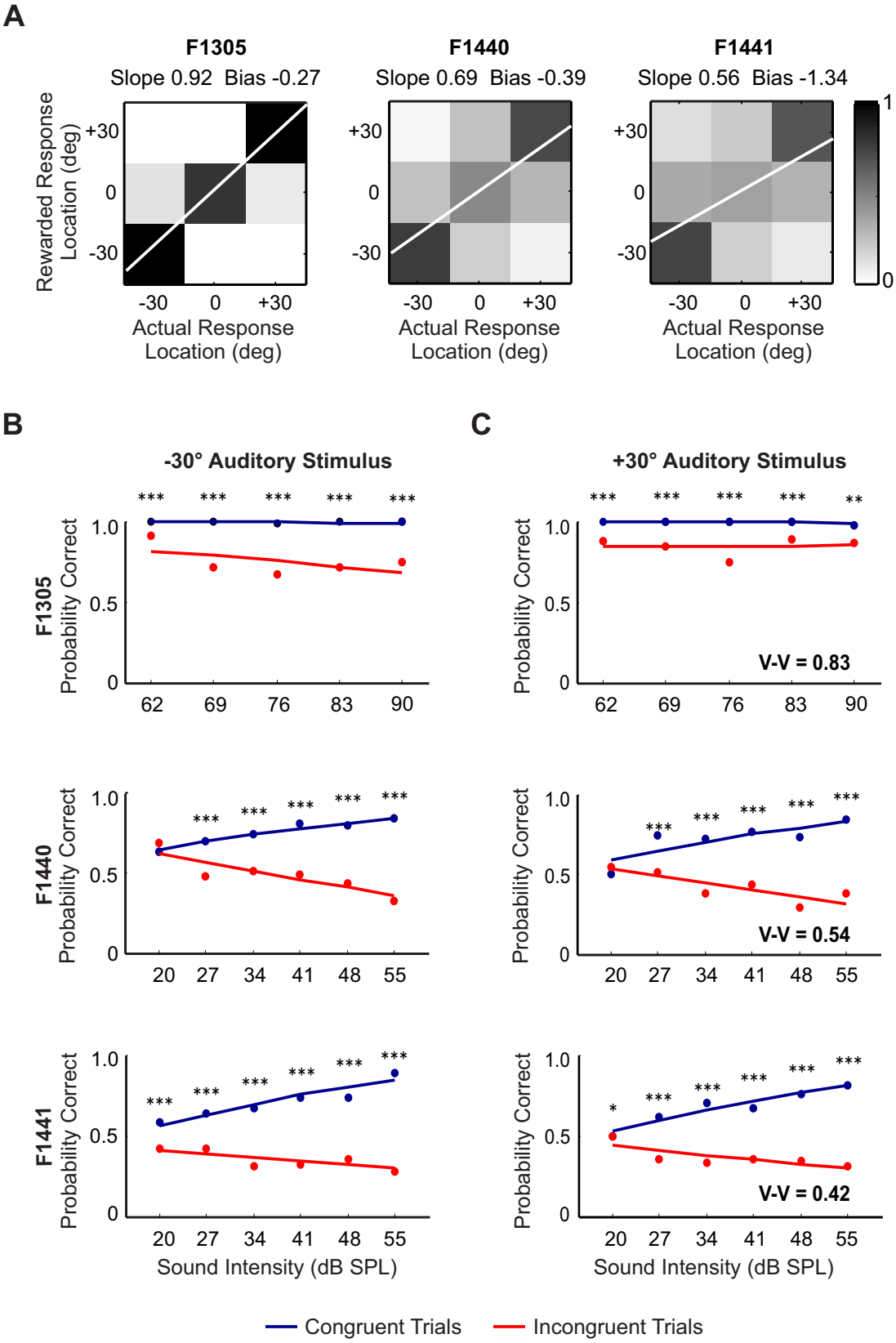


Figure 6-1. Comparison of approach-to-target performance for spatially congruent and incongruent auditory-visual stimuli during the neural recordings. (A) Confusion

matrices illustrating the conditional probability (represented in grey scale) of the actual response location approached for each of the rewarded response locations. Data for individual animals are shown separately and pooled across sound intensities. A regression line is fitted to each confusion matrix, using the linear least squares method, and the corresponding slope (gain) and y-intercept (bias) are shown above each plot. (B-C) Modelled probability of a correct approach-to-target response plotted as a function of sound intensity (F1305: 69 – 90 dB SPL; F1440 & F1441: 20 – 55 dB SPL) for spatially congruent (blue) and incongruent (red) stimulus targets. The model results are shown in separate panels for each animal (rows) and according to whether the auditory stimulus component was presented from -30° (B) or +30° (C). Significant differences between congruent and incongruent trials are indicated by the asterisks and *P*-values by their number (* $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$). For comparison, the probability of a correct response to the fixed brightness visual-visual (V-V, [-30°V +30°V]) stimulus is shown for each animal.

6.2.2 Basic Response Properties of Neurons

A total of 237 units that exhibited an excitatory response to either the auditory and/or visual stimulus were recorded from during the categorisation task and were confirmed, by histological analysis, to reside within early auditory cortical processing areas (Appendix 1). Using the response classification derived from the localisation task, 192 units were classified as auditory, 15 as visual, 23 as bimodal suprathreshold and 7 as bimodal subthreshold. Overall, auditory response thresholds were comparable between tasks (cat vs. loc: 34.7 ± 5.7 vs. 36.2 ± 2.3 dB SPL).

6.2.3 Effect of Stimulus Parameters on Spiking Activity

In order to investigate the effect of varying stimulus properties on evoked activity, scatter plots comparing the number of spikes evoked in a 150 ms window after stimulus onset were generated and a generalised linear model for a negative binomial

distribution was fitted to the spike count data. Analysis of the localisation task recordings illustrated that units typically exhibited a contralateral preference for the auditory stimulus. Likewise, this was also the case for the categorisation task, as illustrated by Figure 6-2A-B, which shows that presentation of the auditory stimulus component from the contralateral hemifield evoked a stronger response than ipsilateral presentation regardless of the stimulus congruency. Furthermore, this result was significant for both congruent (Ipsi vs. Contra auditory units: 0.73 ± 0.08 vs. 0.86 ± 0.07 spikes/trial; bimodal units: 0.74 ± 0.06 vs. 0.86 ± 0.07 spikes/trial) and incongruent stimuli (Ipsi vs. Contra auditory units: 0.69 ± 0.07 vs. 0.83 ± 0.08 spikes/trial; bimodal units: 0.70 ± 0.06 vs. 0.83 ± 0.07 spikes/trial) across most sound intensities (see Fig. 6-2C-D and appendix 6.4.1.2 for significance values).

Although the integrative properties of neurons differ between the cortex and SC, in terms of the magnitude of the response change when unisensory stimuli are combined (Murray & Wallace, 2012), neurons in both regions are still thought to integrate multisensory information according to the spatio-temporal coincidence and effectiveness of stimuli (Wallace *et al.*, 1992). In Chapter 4, we reported the finding that the addition of a spatially congruent visual stimulus had a significant suppressive effect on the average auditory response of the population. Therefore, if the rules that govern multisensory integration in the SC extrapolate to the auditory cortex then it might be expected that the responses evoked by incongruent stimuli would be less suppressed than those evoked by congruent stimuli (i.e. less modulated). However, we found instead that the presentation of the spatially incongruent visual stimulus supressed the overall evoked activity by 4.5%. In view of the contralateral preference of units identified above, the significance of this effect was investigated seperately,

according to whether the auditory stimulus component was presented from the contra- (Fig. 6-2E) (congruent vs. incongruent: auditory units: 0.86 ± 0.08 vs. 0.83 ± 0.08 spikes/trial; bimodal units: 0.87 ± 0.07 vs. 0.83 ± 0.07 spikes/trial) or ipsilateral (Fig. 6-2F) (congruent vs. incongruent: auditory units: 0.73 ± 0.07 vs. 0.68 ± 0.07 spikes/trial; bimodal units: 0.74 ± 0.07 vs. 0.69 ± 0.06 spikes/trial) hemifield relative to each unit. This analysis showed that the activity evoked by the incongruent stimulus was significantly lower than that evoked by the congruent stimulus. However, this effect was only significant when the auditory component of the stimulus was presented from the contralateral hemifield at the lowest sound intensities (see Fig. 6-2G-H and Appendix 6.4.1.2 for significance values). Furthermore, this effect was observed at the lowest sound intensities regardless of whether units were enhanced (significance level range across 20 - 34 dB SPL: t_{2542} = -2.32 - -2.17, P = 0.02 – 0.03) or depressed (significance level range across 20 - 48 dB SPL: t = -3.46 - -2.17, P = <0.001 – 0.02) by multisensory stimulation (see Appendix 6.4.1.3 for all significance values).

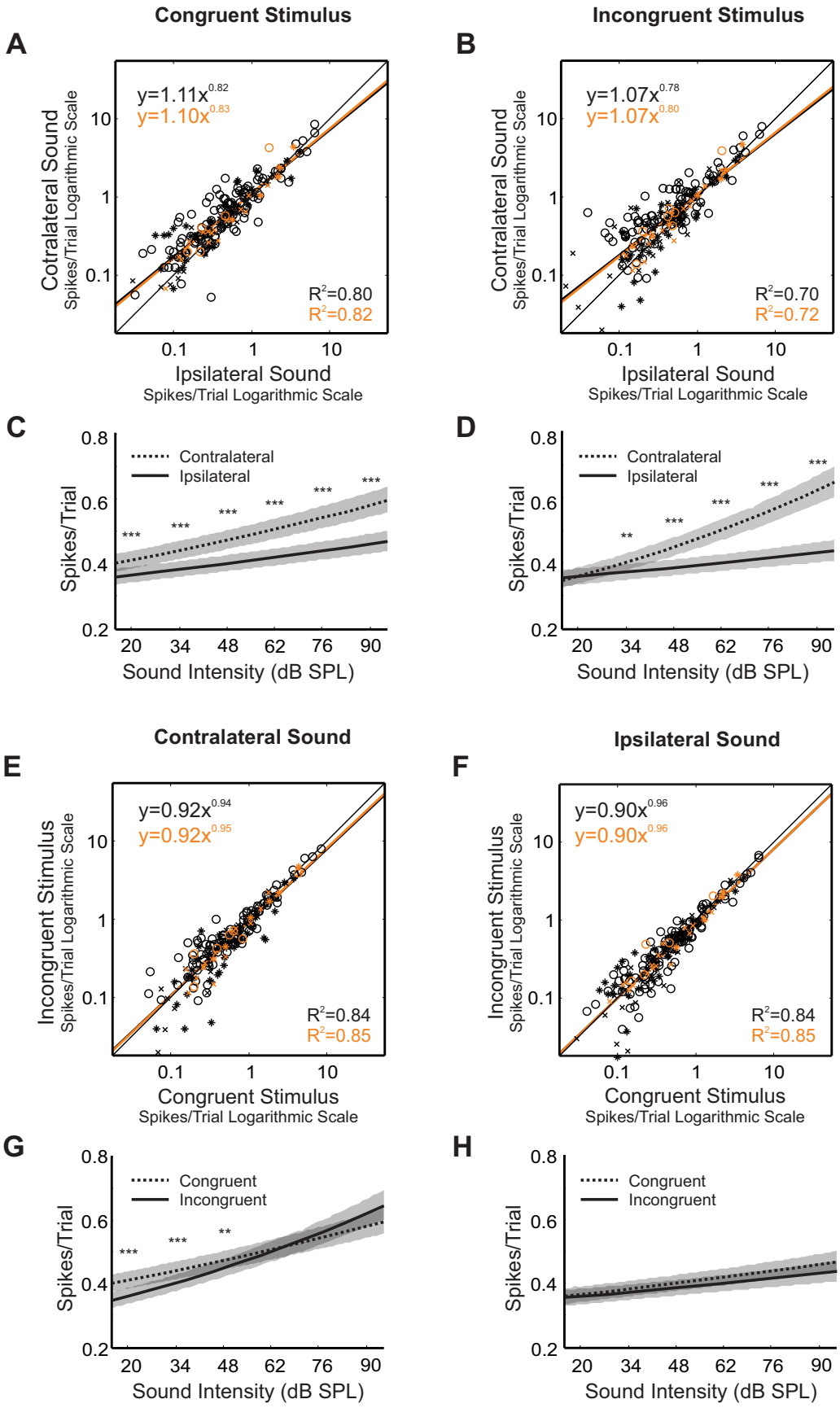

A
Congruent Stimulus
y=1.11x^0.82
y=1.10x^0.83
R²=0.80
R²=0.82
Cotralateral Sound
Spikes/Trial Logarithmic Scale
Ipsilateral Sound
Spikes/Trial Logarithmic Scale
B
Incongruent Stimulus
y=1.07x^0.78
y=1.07x^0.80
R²=0.70
R²=0.72
Contralateral Sound
Spikes/Trial Logarithmic Scale
Ipsilateral Sound
Spikes/Trial Logarithmic Scale
C
Contralateral
Ipsilateral
Spikes/Trial
Sound Intensity (dB SPL)
D
Contralateral
Ipsilateral
Spikes/Trial
Sound Intensity (dB SPL)
E
Contralateral Sound
y=0.92x^0.94
y=0.92x^0.95
R²=0.84
R²=0.85
Incongruent Stimulus
Spikes/Trial Logarithmic Scale
Congruent Stimulus
Spikes/Trial Logarithmic Scale
F
Ipsilateral Sound
y=0.90x^0.96
y=0.90x^0.96
R²=0.84
R²=0.85
Incongruent Stimulus
Spikes/Trial Logarithmic Scale
Congruent Stimulus
Spikes/Trial Logarithmic Scale
G
Congruent
Incongruent
Spikes/Trial
Sound Intensity (dB SPL)
H
Congruent
Incongruent
Spikes/Trial
Sound Intensity (dB SPL)

Figure 6-2. Effect of varying stimulus parameters on the number of spikes evoked per trial. Throughout the figure, units classified as unimodal auditory are denoted in black and units classified as bimodal in orange. (A-B) Scatter plots comparing the number of spikes evoked in a 150 ms window following stimulus onset by contra- and ipsilateral auditory stimulus components. Data are shown separately for congruent (A) and incongruent (B) stimuli. Points above the unity line indicate that units responded more strongly to stimuli when the auditory component was presented from the contralateral hemifield. A regression line was fitted to each scatter plot, using the linear least squares method, and the corresponding slope and y-intercept are shown along with the R^2 value in each panel. Data from individual animals are represented by the different symbols. (C-D) Corresponding modelled spike counts are shown as a function of sound intensity for contralateral (dotted line) and ipsilateral (black line) auditory stimulus components below each scatter plot. Significant differences between stimulus conditions are indicated by the asterisks and probability values by their number (* $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$). (E-F) Scatter plots comparing the number of spikes evoked (0 – 150 ms) by congruent and incongruent stimuli. Data are shown separately according to whether the auditory component of the stimulus was presented from the contra- (E) or ipsilateral (F) hemifield relative to each unit. (G-H) Corresponding modelled spike count values are shown as a function of sound intensity for congruent (dotted line) and incongruent (black line) stimuli below each scatter plot (significance levels as above).

In addition to firing rate, the effect of stimulus congruency on the timing and variability of spike discharge patterns was investigated by calculating response latencies and Fano Factor values (see Methods 2.6.5 and Chapter 4.2.4), respectively. If the neural activity reflects the behavioural outcomes then it might be expected that congruent stimuli evoke more reliable responses with shorter latencies.

Inspection of the post stimulus time histogram for data pooled across sound intensities, however, shows that the temporal response dynamics of neurons were similar for congruent and incongruent stimuli (Fig. 6-3A). To investigate whether a significant difference exists between stimulus conditions, a linear model was fitted to the mean latency values (Fig. 6-3B). This analysis showed that regardless of stimulus congruency, response latencies were significantly shorter at higher sound intensities (mean latencies: congruent: $t_{2542} = -5.68$, $P < 0.001$; incongruent: $t_{2542} = -4.65$, $P < 0.001$). Overall, latencies were marginally shorter for congruent stimuli than for

incongruent stimuli (mean: congruent vs. incongruent: 56.7 ± 0.92 vs. 58.1 ± 0.97 ms). However, this effect was not significant for the majority of stimulus intensities (see Fig. 6-3B and Appendix 6.4.1.4 for significance values).

Overall, Fano Factor (FF) values were higher for congruent than incongruent stimuli, suggesting that the neural responses evoked by the former are less reliable than those evoked by the latter (congruent vs. incongruent: 1.98 ± 0.07 vs. 1.92 ± 0.06). This trend was observed regardless of whether units were enhanced (congruent vs. incongruent: 2.06 ± 0.10 vs. 1.99 ± 0.09) or depressed (congruent vs. incongruent: 1.87 ± 0.10 vs. 1.81 ± 0.09) by multisensory stimulation. However, this effect was not significant across sound intensities. (see Fig. 6-3C and Appendix 6.4.1.4 for significance values). In addition, no effect of sound intensity on FF values was found (congruent: $t_{1824} = -0.12$, $P = 0.90$; incongruent: $t_{1824} = -1.95$, $P = 0.05$).

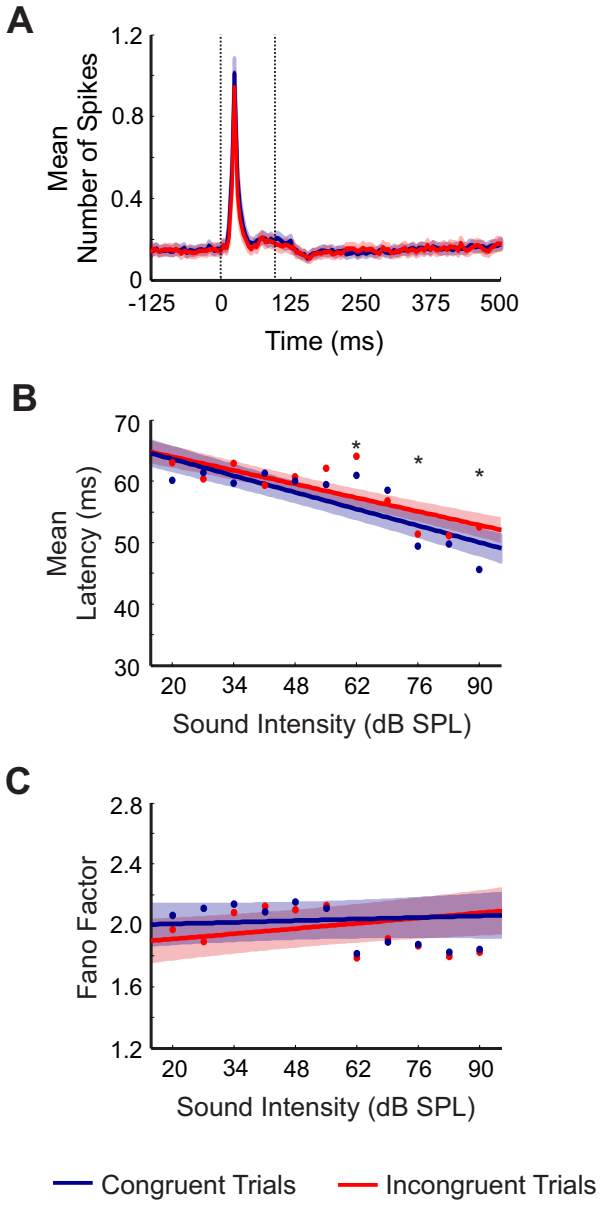


Figure 6-3. Comparison of spike timing and variability for spatially congruent (blue) and incongruent (red) stimuli. (A) Pooled peri stimulus time histogram showing the mean number of spikes evoked (in 5 ms time bins) over time relative to stimulus onset. (B-C) Modelled mean spike latencies (B) and Fano Factor values (C) shown as a function of sound intensity together with the actual values (filled circles). Significant differences between stimulus conditions are indicated by the asterisks and probability values by their number (* $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$).

6.2.4 Correlation of Neural Activity and Behavioural Outcomes

In order to investigate the relationship between evoked activity and trial correctness, histograms illustrating the spike rate relative to the stimulus onset and scatter plots comparing the number of spikes evoked in a 150 ms window after stimulus onset were generated seperately for congruent (Fig. 6-4) and incongruent (Fig. 6-5) trials.

For congruent trials, inspection of the histograms shows that, regardless of whether stimuli were presented from the contra- or ipsilateral hemifield, the peak spike rate was higher for correct approach-to-target responses than for incorrect responses (Fig. 6-4A-B) (correct vs. incorrect peak spike rate: contralateral: 9.9 vs. 7.4 s^{-1}; ispsilateral: 7.2 vs. 6.2 s^{-1}), which is consistent with our results for equivalent auditory-viusal stimuli from the localisation task. Overall, the number of spikes in a 150 ms window after stimulus onset was marginally higher for correct responses than for incorrect responses (Fig. 6-4C-D) (correct vs. incorrect: contralateral: 0.88 ± 0.08 vs. 0.83 ± 0.08 spikes/trial; ipsilateral: 0.75 ± 0.08 vs. 0.73 ± 0.08 spikes/trial). However, a generalised linear mixed effects model fitted to the spike count data found that this effect was not significant across stimulus locations and sound intensities (see Fig. 6-4E-F and Appendix 6.4.1.5 for significance values).

For incongruent stimuli with the auditory stimulus component presented from the contralateral hemifield, the peak spike rate (Fig. 6-5A) (correct vs. incorrect: 8.3 vs. 8.2 s^{-1}) and number of spikes evoked in a 150 ms window after stimulus onset (Fig. 6-5C) (correct vs. incorrect: 0.83 ± 0.07 vs. 0.83 ± 0.06 spikes/trial) were comparable regardless of the behavioural outcome. However, for incongruent stimuli with the auditory stimulus component presented from the ipsilateral hemifield, the

peak spike rate (Fig. 6-5B) (correct vs. incorrect: 5.9 vs. 7.2 s^{-1}) and number of spikes evoked (Fig. 6-5D) (correct vs. incorrect: 0.67 ± 0.07 vs. 0.71 ± 0.07 spikes/trial) were slightly higher for incorrect approach-to-target responses than for correct approach-to-target responses. For the incongruent stimulus, a generalised linear model fitted to the spike count data found no significant effect of trial correctness on evoked activity across all stimulus locations and sound intensities (see Fig. 6-4E-F and Appendix 6.4.1.5 for significance values). In summary, our results suggest that neural activity in the auditory cortex during the categorisation task does not significantly correlate with behavioural outcomes.

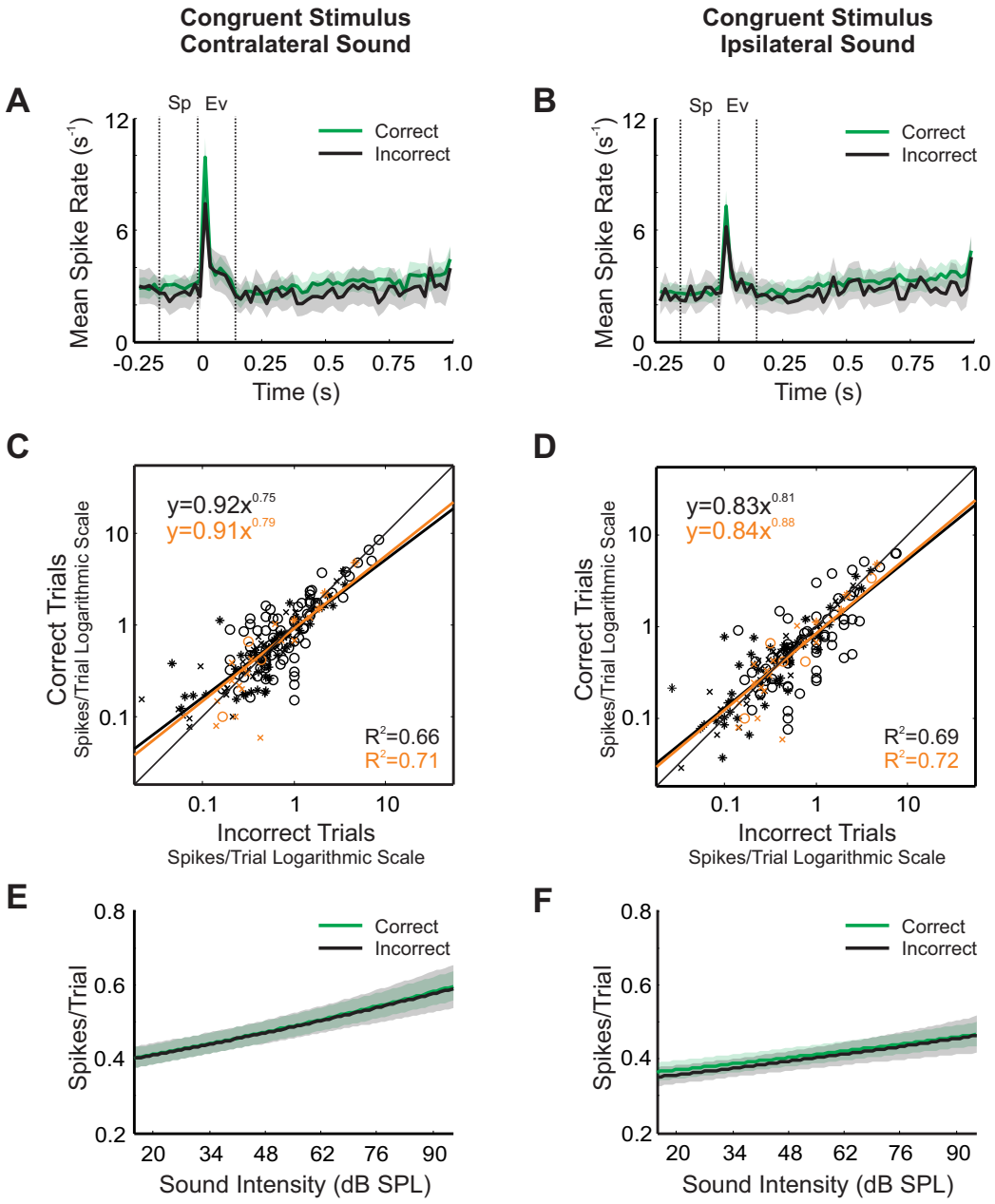


Figure 6-4. Relationship between evoked activity and response correctness for congruent trials. For simplicity data are shown separately throughout the figure according to whether stimuli were presented from the contra- (left panels) or ipsilateral (right panels) hemifield relative to each unit. (A-B) Pooled histograms showing the mean spike rate (in 5 ms bins) over time relative to the stimulus onset for correct (black lines) and incorrect trials (green lines). (C-D) Scatter plots comparing the number of spikes evoked in a 150 ms window following stimulus onset for correct and incorrect trials (auditory units, black; bimodal units, orange). A regression line was fitted to each scatter plot, using the linear least squares method, and the corresponding slope and y-intercept are shown along with the R^2 value in each panel. Data from individual animals are represented by the different symbols. (E-F) Corresponding modelled spike counts are shown as a function of sound intensity for correct (green line) and incorrect (black line) trials. No significant difference was found between trial types.

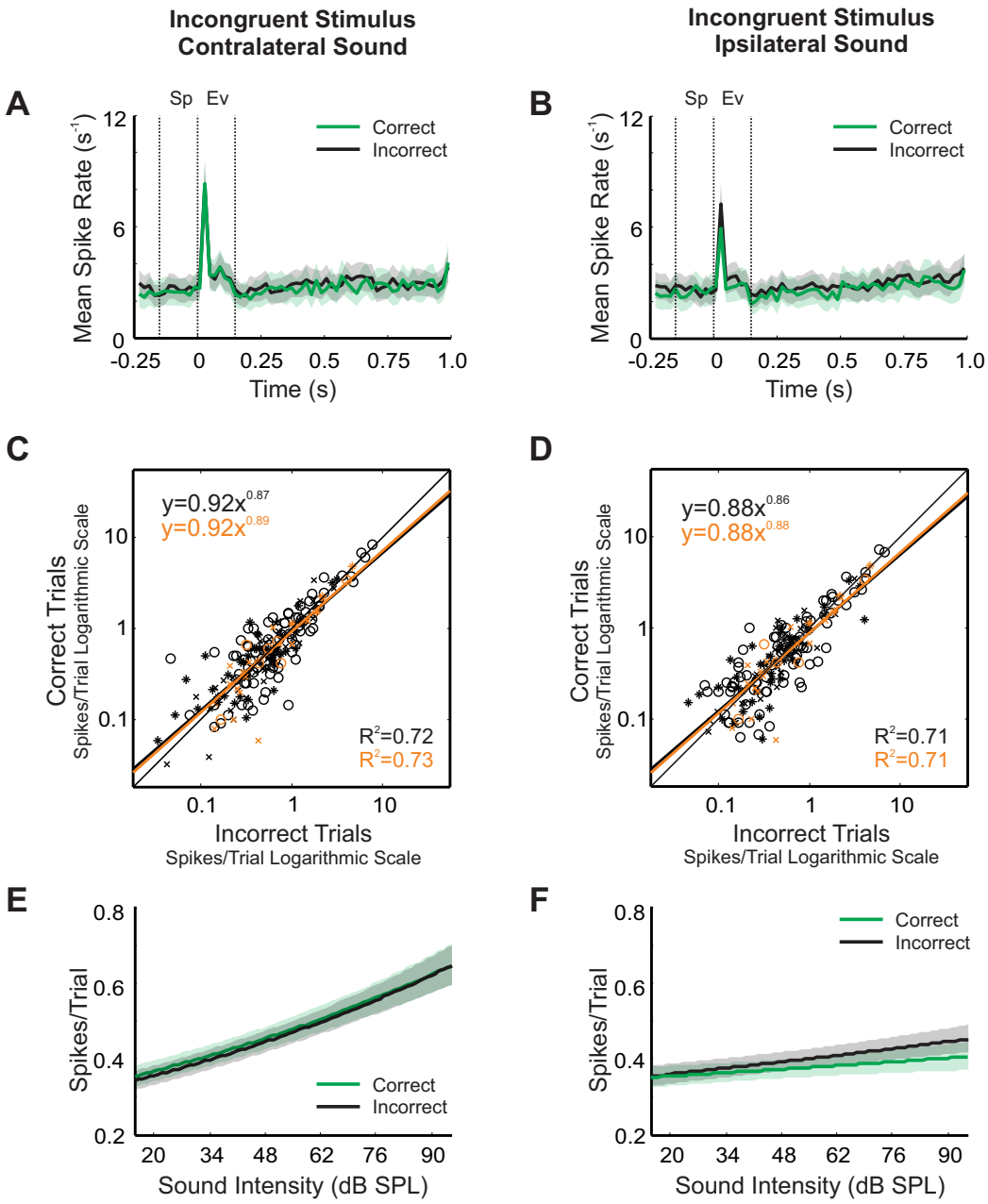


Figure 6-5. Relationship between evoked activity and response correctness for incongruent trials. For simplicity data are shown separately according to whether the auditory stimulus component was presented from the contra- (left panels) or ipsilateral (right panels) hemifield relative to each unit. (A-B) Pooled histograms showing the mean spike rate (in 5 ms bins) over time relative to the stimulus onset for correct (black lines) and incorrect trials (green lines). (C-D) Scatter plots comparing the number of spikes evoked in a 150 ms window following stimulus onset for correct and incorrect trials (auditory units, black; bimodal units, orange). A regression line was fitted to each scatter plot, using the linear least squares method, and the corresponding slope and y-intercept are shown along with the R^2 value in each panel. Data from individual animals are represented by the different symbols. (E-F) Corresponding modelled spike counts are shown as a function of sound intensity for correct (green line) and incorrect (black line) trials. No significant difference was found between trial types.

The analysis of pre-implantation error data from incorrect incongruent trials provided an insight into the interplay of attention capture between the modalities during task performance and showed that animals overwhelmingly oriented towards and then approached the auditory rather than the visual stimulus component. Likewise, during the neural recordings, all animals predominantly approached the auditory rather than the visual stimulus component (Fig. 6-6A). In order to explore whether the neural activity from incorrect incongruent trials reflected this behavioural outcome, scatter plots comparing the evoked activity for auditory versus visual approach-to-target responses were generated and a generalised linear model was fitted to the spike count data (Fig. 6-6B-C). On average, evoked activity was higher for auditory than for visual approach-to-target responses regardless of whether the auditory stimulus component was presented from the contra- (auditory vs. visual response auditory units: 0.81 ± 0.08 vs. 0.70 ± 0.09 spikes/trial; bimodal units: 0.82 ± 0.08 vs. 0.74 ± 0.09 spikes/trial) or ipsilateral (auditory vs. visual response auditory units: 0.71 ± 0.08 vs. 0.69 ± 0.09 spikes/trial; bimodal units: 0.73 ± 0.07 vs. 0.68 ± 0.08 spikes/trial) hemifield. However, this effect was not significant across the range of sound intensities tested (Fig. 6-6D-E) (range of significance values across intensities contralateral: t_{13396} = -1.44 – -0.42, P = 0.15 – 0.67; ipsilateral: t_{13396} = -0.07 – 0.26, P = 0.79 – 0.94). The behavioural bias observed towards the auditory stimulus component could be explained by a lack of detection of the visual stimulus. The finding that the presentation of a spatially incongruent visual stimulus resulted in a subtle but significant suppression of the evoked activity, suggests that at the neural level there is a signature of its register. Together, these results suggest that on incorrect incongruent trials the animal might fail to register the visual stimulus.

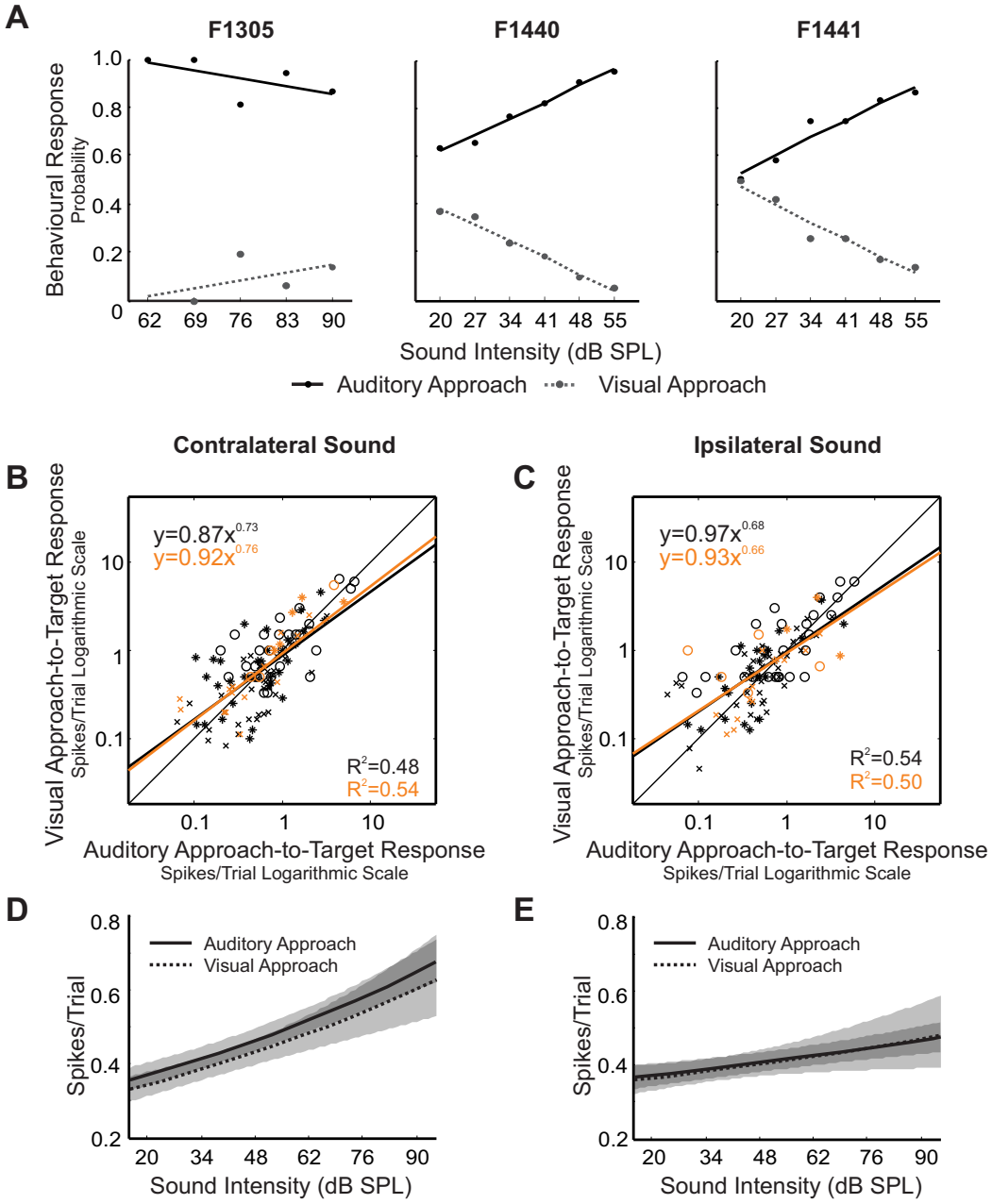


Figure 6-6. Behavioural and neural responses on incorrect incongruent trials. (A) Modelled probability of an incorrect approach-to-target response to the auditory (black solid line) or visual (grey dotted line) stimulus component shown, together with the actual probability values (solid dots), as a function of sound intensity. Data for individual animals are shown in separate panels. (B-C) Scatter plots comparing the number of spikes evoked in a 150 ms window following stimulus presentation depending on the stimulus modality approached (auditory units, black; bimodal units, orange). Data are shown separately according to whether the auditory stimulus component was presented from the contra- (left panel) or ipsilateral (right panel) hemifield relative to each unit. A regression line was fitted to each scatter plot, using the linear least squares method, and the corresponding slope and y-intercept are shown along with the R^2 value in each panel. Data from individual animals are represented by the different symbols. (D-E) Corresponding modelled spike counts are shown as a

function of sound intensity for auditory (solid black line) and visual (grey dashed line) approach-to-target responses below each scatter plot. No significant difference in evoked activity was found between auditory and visual approach-to-target responses.

The use of awake-behaving recordings enabled us to investigate whether the evoked activity is more closely correlated with the stimulus characteristics or the animal's percept, as indicated by its behavioural choice, using a population decoding analysis to the spiking activity (Meyers, 2013) (see Section 2.6.5). The pattern classifier was used to predict the experimental condition (i.e. sound intensity, sound location, stimulus congruency and trial correctness) on a particular trial (Fig. 6-7A-D). The results of this analysis for sound intensity are shown in Figure 6-7A (top panel), which shows that prior to stimulus onset the classification accuracy is at chance level (i.e. ~ 20%) but, shortly after stimulus onset, the classification accuracy peaks at ~ 40%. In order to investigate whether information is encoded by a dynamic population code, a temporal cross training matrix is also plotted (Fig. 6-7A, bottom panel), which illustrates the classification accuracy when the pattern classifier is trained at one time point and tested at a different time point. The results show that the highest decoding accuracies occur along a diagonal, indicating that information is encoded by a dynamic population code. The results for sound location are shown in Figure 6-7B, which shows that the pattern classifier can also be used to predict whether the auditory stimulus component was presented from the ipsi- or contralateral hemifield relative to the unit, given that shortly after stimulus onset the classification accuracy increases from chance level (~ 50%) to peak at ~ 80%. In contrast, the results for stimulus congruency and trial correctness shown in Figures 6-7C-D, respectively, show that the classifier could not predict the stimulus

congruency or whether the animal made a correct or incorrect approach-to-target response. In summary, the population decoding analysis suggests that the activity in the early auditory processing areas is informative about the physical properties of the auditory stimulus rather than the cross-modal binding of stimuli or the animal's perceptual choice.

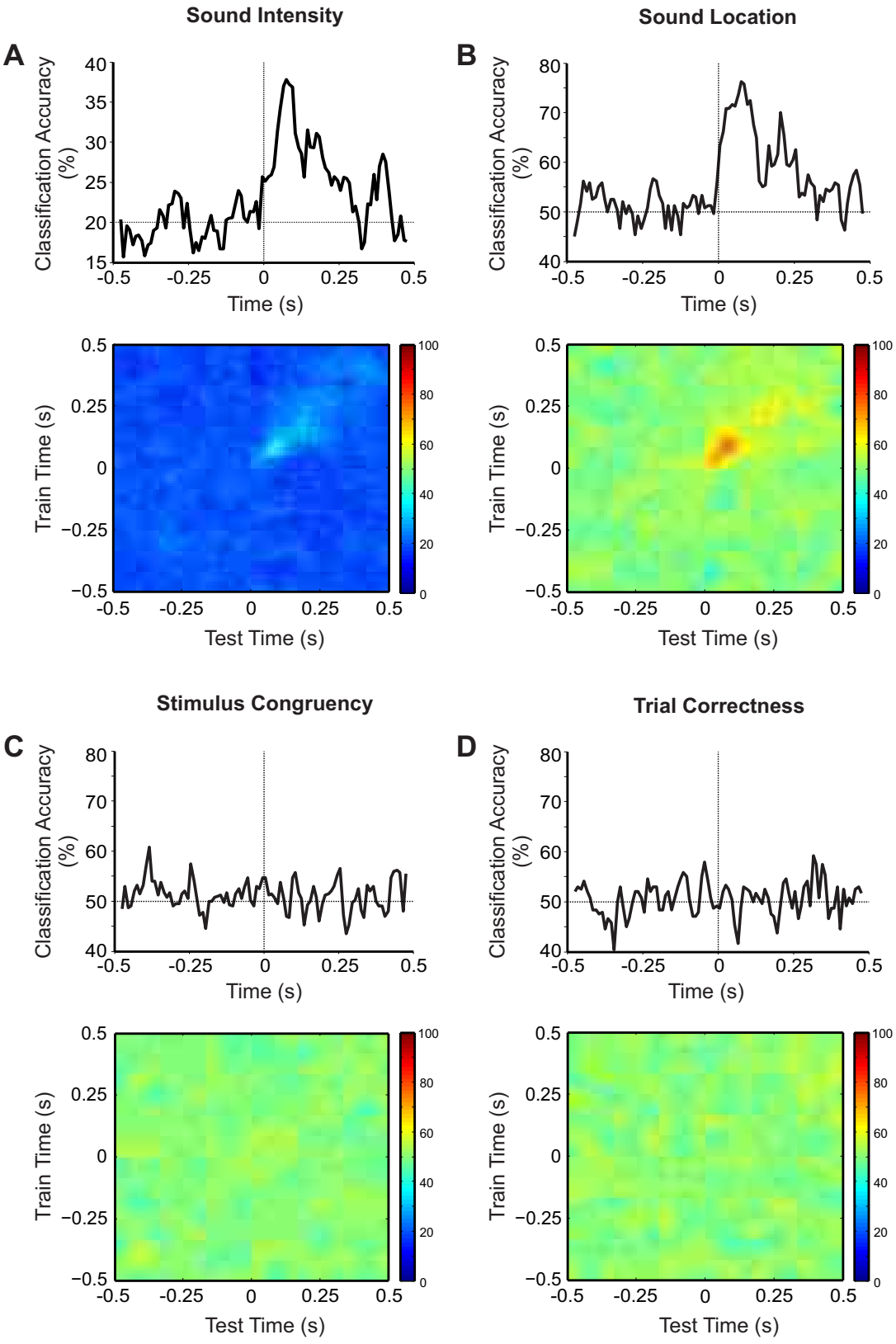


Figure 6-6. Results of the population decoding analysis. (A-D, top panels) Accuracy of the pattern classifier shown as a percentage in decoding sound intensity (A), sound

location (B), stimulus congruency (C) and behavioural response correctness (D) from neural activity in early auditory cortical areas. The black dashed vertical line shows the stimulus onset and the black dashed horizontal line shows the decoding accuracy at chance level. For example, Figure A (top panel) shows that prior to stimulus onset the decoding accuracy is at approximately chance level (20%) and rises quickly after stimulus onset to peak at 37%. (A-D, bottom panels) Corresponding temporal cross-training matrices are shown below each plot illustrating the classification accuracy as a percentage (on a colour scale) when the pattern classifier is trained at one time point (y-axis) and tested at a different time point (x-axis). For example, the results shown in Figure B (bottom panel) show that information is contained in a dynamic population code, which can be seen by the fact that the highest decoding accuracies occur along the diagonal of the plot.

6.2.5 Comparison of Neural Activity between Behavioural Tasks

Finally, we examined whether the context of the behavioural task affects neural activity by comparing the activity recorded from equivalent spatially congruent auditory-visual trials from the localisation and categorisation tasks (i.e. [-30°V, -30°A] or [+30°V, +30°A]). For animals F1440 and F1441, approach-to-target performance was better for the localisation task, whereas for animal F1305 approach-to-target performance was better for the categorisation task (Fig. 6-8A). However, for all animals, evoked activity was higher for the localisation task than for the categorisation task, regardless of whether stimuli where presented from the contralateral (Fig. 6-8B) (loc vs. cat: F1305: 1.07 vs. 0.97; F1440: 1.13 vs. 0.84; F1441: 0.85 vs. 0.65 spikes/trial) or ipsilateral (Fig. 6-8C) (loc vs. cat: F1305: 0.84 vs. 0.82; F1440: 0.91 vs. 0.76; F1441: 0.66 vs. 0.58 spikes/trial) hemifield relative to each unit. Furthermore, a generalised linear model fitted to the spike count data showed that this effect was significant across all sound intensities (see Fig. 6-8D-E and Appendix 6.4.1.6 for significance values). It is unlikely that this difference arises because of the task design, given that the order of the behavioural tasks was reversed

every recording cycle and, in addition, spontaneous activity levels were comparable between tasks (loc vs. cat: contralateral: 0.47 ± 0.05 vs. 0.50 ± 0.05; ipsilateral: 0.48 ± 0.04 vs. 0.46 ± 0.05 spikes/trial). A population decoding analysis was performed in order to explore this effect further by investigating whether the task type (i.e. localisation vs. categorisation) could be predicted from the spiking activity (Fig. 6-8F-G). This analysis suggests that the activity in the auditory cortex is informative about the task type, given that shortly after stimulus onset the classification accuracy increases from chance level (~ 50%) to peak at ~ 70%.

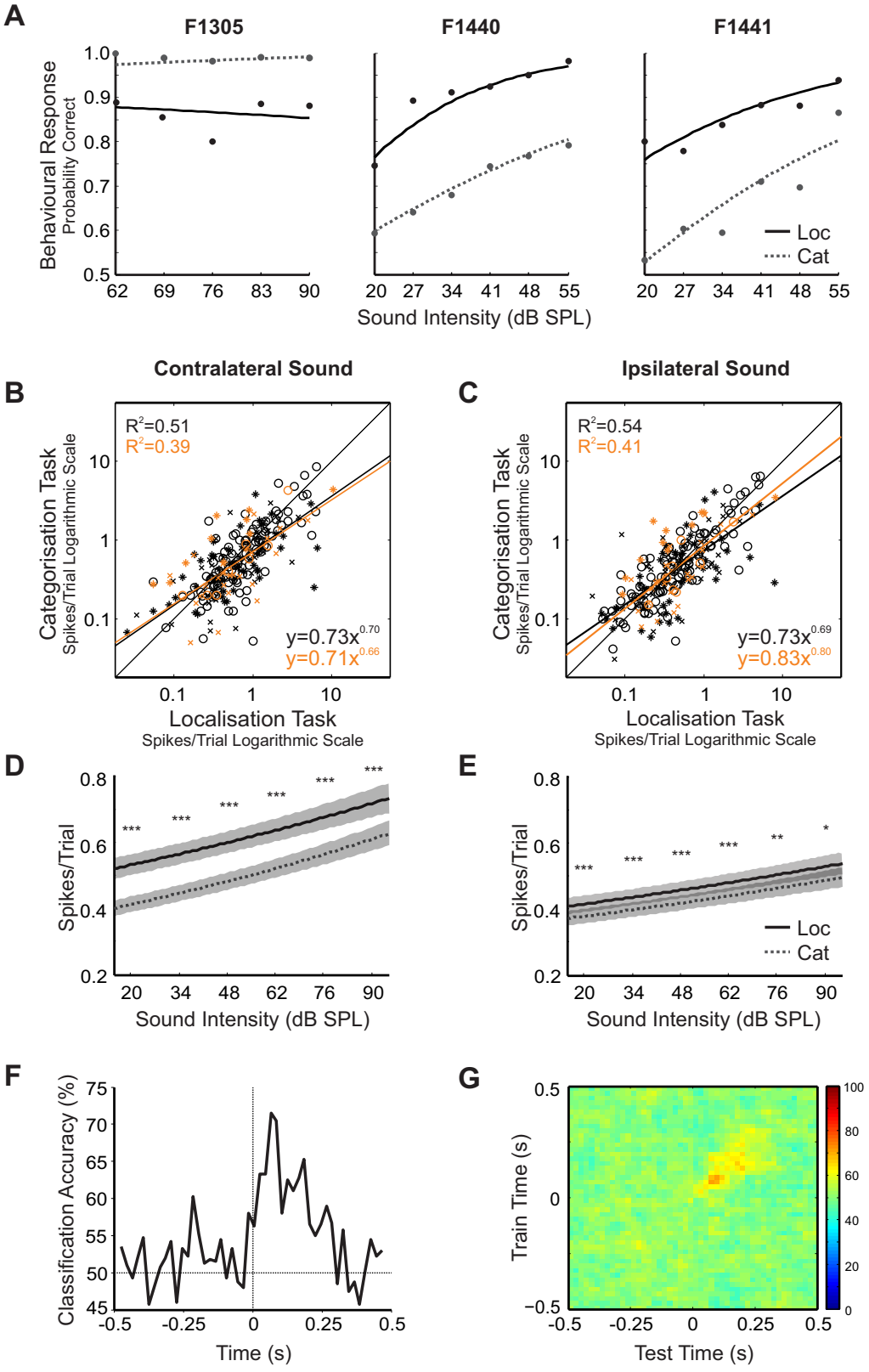


Figure 6-8. Comparison of behavioural performance and neural activity between the categorisation and localisation tasks. For both tasks, only data from spatially congruent

auditory-visual trials were included. (A) Modelled probability of a correct approach-to-target response is shown, together with actual probabilities (filled circles), as a function of sound intensity for the localisation (solid black line) and categorisation tasks (dashed grey line). Data from individual animals are shown in the separate panels. (B-C) Scatter plots comparing the number of spikes evoked in a 150 ms window following presentation of categorisation and localisation task stimuli (auditory units, black; bimodal units, orange). A regression line was fitted to each scatter plot, using the linear least squares method, and the corresponding slope and y-intercept are shown along with the R^2 value in each panel. Data from individual animals are represented by the different symbols. (D-E) Corresponding modelled spike counts are shown as a function of sound intensity for the localisation (solid black line) and categorisation (dashed grey line) tasks below each scatter plot. Significant differences between responses are indicated by the asterisks and probability values by their number (* $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$). (F-G) Population decoder classification accuracy results shown as a function of time relative to stimulus onset (F) and corresponding temporal cross-training matrix (G) from decoding task type (i.e. categorisation vs. localisation).

6.3 Discussion

In this chapter, neural activity in the auditory cortex was explored whilst animals categorised the spatial congruency of auditory-visual stimuli. As for the pre-implantation behavioural task results, we found that animals were able to categorise temporally coincident auditory-visual stimuli as one object in the spatially congruent stimulus condition or as discrete objects in the incongruent condition. Understanding how spatial information across these two sensory modalities is combined is challenging because visual and auditory signals are encoded in different ways within their respective sense organs and brain pathways. Positions in visual space are represented by the locus of activity within topographic projections from the retina, whereas the auditory system is organised tonotopically, which means that the location of a sound source has to be derived from monaural and binaural spatial cues that arise from the geometry of the head and external ears (King *et al.*, 2001).

Our localisation task results suggest that rather than driving activity, visual inputs to the auditory cortex predominantly have a subthreshold modulatory effect on the responsiveness of the cortical network. Therefore, it is possible that this effect subserves functions that would not necessarily require visual inputs to the auditory cortex to be in topographic alignment with auditory representations including, for example, enhancing stimulus detection, degrading distracting sensory processing in neighbouring cortices or re-directing attention. In support of this hypothesis, our findings from the localisation task suggest that the auditory cortex does not act as a substrate for visual influences on sound localisation behaviour. However, our findings from the localisation task also show that sound source location can be

represented by a hemispheric population code, which could provide an effective means for auditory and visual spatial information to be combined within the auditory cortex. For example, we found that activity in the cortex contralateral to the auditory stimulus was suppressed by incongruent stimuli, whereas activity in the ipsilateral cortex was unaffected, thus reducing the difference in activity between the left and right auditory cortical areas. Although, not directly tested in this experiment, it is feasible that a direct relationship between the magnitude of this modulatory effect and the spatial disparity of stimuli exists.

In support of this hypothesis, visual influences on auditory cortical activity have been proposed to provide a reference for calibrating the perception of auditory space. For example, ventriloquism and related phenomena (Zwiers et al., 2003) illustrate that vision can alter the perceived location of sound sources in order to resolve short-term spatial conflicts between these modalities. The neural basis for the ventriloquism illusion is poorly understood, but event-related potential and fMRI measurements have revealed changes in activity levels in the auditory cortex on trials in which participants experienced a shift in perceived sound location in the direction of a misaligned visual stimulus (Bonath *et al.*, 2007, 2014). In humans, the ventriloquism illusion breaks down after ~ 15° (Slutsky & Recanzone, 2001). Therefore, although it is highly unlikely that a ventriloquism percept was evoked in the present study since the spatial discrepancy between stimuli was 60°, our results showed an analogous change in activity levels between hemispheres.

The presentation of auditory and visual stimuli from opposite hemifields would have likely evoked activity in opposite hemispheres. Although it is becoming increasingly clear that multiple circuits exist for mediating the influence of one

modality on another, as discussed in Chapter 4, the activation of one primary sensory cortex has recently been shown in rodents to have an inhibitory effect on neighbouring ipsilateral primary sensory cortices by altering the state of the local network via intra-hemispheric cortico-cortical connections (Iurilli *et al.*, 2012; Stehberg *et al.*, 2014). From an ethological perspective, it would seem reasonable to suggest that competition between simultaneous events occurring within separate hemifields would be greater than competition between events occurring within the same hemifield. Although not established, it is possible that inhibitory inter- as well as intra-hemispheric connections exist between different sensory cortices, in which case our finding that the presentation of the incongruent stimulus resulted in an increased suppression of the mean auditory response in the contralateral hemisphere might reflect the relative density of projections within and across hemispheres.

Overall, the presentation of the spatially incongruent stimulus as compared to the congruent stimulus suppressed the evoked activity by 4.5%. A population decoding analysis showed that this modulation was insufficient to predict the trial type from the spike count data. In contrast, the spike count data was highly informative about the physical properties of the auditory stimulus. In addition to encoding information about the physical properties of sound, auditory cortical activity has been shown to be informative about the animal's impending sensory decision (Selezneva *et al.*, 2006; Niwa *et al.*, 2012b; Bizley *et al.*, 2013; Tsunada *et al.*, 2015). For example, spiking activity recorded from auditory cortex, including from primary areas, of ferrets trained to discriminate the pitch of an artificial vowel has been shown to encode both the fundamental frequency of the target sound as well as the perceived pitch, with choice-related activity increasing just before the

behavioural response (Bizley *et al.*, 2013). In contrast, in the present study we found no evidence for decision-related activity in the early auditory cortex. It seems likely that this discrepancy in findings may arise in part because of fundamental differences in the respective tasks, both in terms of the task design and the level of abstraction required to form the decision. For example, the aforementioned studies used unimodal tasks and multiple stimulus presentations per trial, whereas the present study employs a cross-modal task with a single stimulus presentation per trial, which is equally reliant on the use of visual as well as auditory information to form the decision. In Chapter 4, we showed that visual influences in the early auditory cortex predominantly have a modulatory effect. Likewise, in the present chapter we have presented results showing that the spatial congruency of the stimulus could not be decoded from the spike count data. Therefore, it is perhaps not surprising that in the present task, choice-related activity was not identified in the auditory cortex since it seems likely there is insufficient visual information available, or integrative capability at this level, to support decision-making in a cross-modal spatial task.

A comparison of the localisation and categorisation task results showed that the activity evoked by equivalent stimulus combinations varied between tasks, whereby the highest level was observed for the localisation task. Furthermore, a population decoding analysis showed that the auditory cortical evoked activity was informative about which task was being performed. Similarly, electrophysiological recordings from the ferret auditory cortex have shown task dependent changes in evoked activity. For example, when tone detection and discrimination tasks were performed sequentially neurons responded differently to identical tones, manifest as a change in their spatio-temporal receptive fields, (Fritz *et al.*, 2005). However, it is

difficult to pinpoint exactly which element of the behavioural context is responsible for the difference in activity observed between tasks.

The behavioural context of the localisation and categorisation tasks depends on the different stimulus-response associations made to all trial types within each task and not only to a subset of them. Therefore, although the trials included in the analysis were identical, in terms of the stimulus composition and rewarded response location, there are clear differences between them in terms of their behavioural context. In order to perform the localisation task, it is plausible that the animals could have used an "approach the stimulus" strategy for all trials, whereas for the categorisation task this strategy would not have worked, given the spatial disparity between the stimulus and response location on a subset of trials. Instead, for the categorisation task, it is possible that the animals adopted a "wait and see" strategy, which could have the advantage of providing additional information resulting, for example, from a longer integration time. In support of this hypothesis, ferrets have previously been shown to use different response strategies, for example a trade-off between speed and performance accuracy, during sound localisation tasks (Leach *et al.*, 2013). In view of the readiness of ferrets to respond during behavioural tasks, illustrated by their quick response times and high false alarm rates in go/no-go tasks (Homma et al., 2016), it is plausible that the "wait and see" strategy requires active suppression of the "ready to go" strategy. This suggestion is in keeping with the paradoxical suppression of sound evoked responses observed in rats when they are engaged with an auditory task as compared to passively listening to identical stimuli (Otazu *et al.*, 2009). Alternatively, responses to visual and tactile stimuli as well as motor-related activity have been shown to emerge during training in the auditory

cortex, which also raises the possibility that the training received by animals is responsible for the emergence of task-specific activity patterns (Brosch *et al.*, 2005).

However, it is possible that other mechanisms also play a role in the observed task related modulation of neural activity. In the cortex, the myriad of synaptic input driven by on-going cortical activity has been shown to modulate the state of local networks and thereby affect neural responsiveness (Tsodyks *et al.*, 1999; Haider & McCormick, 2009). Network state has been shown to be influenced by a number of factors that could have contributed to the difference in activity observed between tasks in this study, including reward expectation (Shuler & Bear, 2006), motivation (Fontanini & Katz, 2006), attention (Lakatos *et al.*, 2008b), motor activity (Wolpert *et al.*, 1995) and decision making by top-down connections (Rodgers & DeWeese, 2014). Although our experimental design cannot extricate the relative contribution of these factors, all could have feasibly contributed to the task-related modulation of neural activity observed in this study. Our results suggest that the importance of the behavioural context is such that it is already reflected in early auditory cortical processing areas. However, it is still unclear whether the observed changes are the result of active optimised encoding processes (i.e. multiplexing) according to the demands of the task or simply a passive reflection of other parallel processes.

6.4 Appendix

6.4.1 Statistical Analyses

6.4.1.1 Approach-to-Target Performance

To investigate the effect of stimulus condition (i.e. congruent vs. incongruent) on approach-to-target response correctness within stimulus intensity and location, a generalised linear mixed effects model for Bernoulli distributed responses was fitted separately to the data from each animal. The R code for a single animal (F1305), sound intensity (20 dB SPL) and sound location (-30°) combination is shown below. Consecutive re-levelling of the model for stimulus location and adjustment of the intercept for sound intensity yielded the results shown in Table 6-1.

```
> mod.F1305 <- glm(Correctness ~ Congruency*Sound_Loc*I(Sound_Level-
20), family = binomial, data = F1305)
```

Intensity	P_{Con}	P_{Incon}	*t*-value	*P*-value
F1305: Sound Location -30°				
62	0.993	0.812	3.668	<0.001
69	0.992	0.783	5.182	<0.001
76	0.990	0.751	7.090	<0.001
83	0.988	0.716	6.633	<0.001
90	0.986	0.679	4.654	<0.001
F1305: Sound Location +30°				

62	0.998	0.838	3.255	0.001
69	0.997	0.841	4.127	<0.001
76	0.995	0.843	5.259	<0.001
83	0.992	0.846	4.871	<0.001
90	0.988	0.849	3.028	0.003
F1440: Sound Location -30°				
20	0.645	0.623	0.447	0.655
27	0.693	0.570	3.369	<0.001
34	0.737	0.516	7.733	<0.001
41	0.777	0.461	10.730	<0.001
48	0.813	0.407	10.761	<0.001
55	0.843	0.356	9.981	<0.001
F1440: Sound Location +30°				
20	0.588	0.534	1.079	0.281
27	0.645	0.488	4.346	<0.001
34	0.699	0.441	8.989	<0.001
41	0.748	0.396	11.468	<0.001
48	0.791	0.352	11.013	<0.001
55	0.828	0.311	10.060	<0.001
F1441: Sound Location -30°				
20	0.555	0.413	3.564	<0.001
27	0.626	0.388	8.064	<0.001
34	0.692	0.363	14.288	<0.001
41	0.75	0.339	17.775	<0.001
48	0.802	0.316	16.803	<0.001
55	0.845	0.294	14.964	<0.001
F1441: Sound Location +30°				
20	0.529	0.435	2.403	0.016
27	0.593	0.406	6.541	<0.001
34	0.654	0.377	12.259	<0.001
41	0.710	0.350	15.026	<0.001
48	0.761	0.32	14.078	<0.001

55	0.805	0.298	12.632	<0.001

Table 6-1. Results of the generalised linear model fitted to approach-to-target data. Modelled probabilities of a correct response for congruent (P_{Con}) and incongruent (P_{Incon}) stimuli are shown for each sound location – intensity combination, together with the *t*- and *P*-values for the following comparison: congruent vs. incongruent.

6.4.1.2 The effect of Stimulus Parameters on Evoked Activity

A negative binomial generalised linear mixed effects model was fitted to the spike count data obtained from the categorisation task. Mixed effects modelling enabled the contribution of fixed effects (i.e. stimulus condition, sound intensity, sound location, and spontaneous count) on the number of spikes evoked per trial (in a 150 ms window after the stimulus onset) to be established, while controlling for random effects (i.e. cell identity nested within ferret identity). Inspection of the data showed that a negative binomial distribution was required since the data were over-dispersed (i.e. the variance exceeded the mean). The model was fitted in R using the code shown below. Consecutive re-levelling of the model for stimulus location and adjustment of the intercept for sound intensity yielded the results shown in Table 6-2.

```
> mod1 <- glmmPQL(formula(paste("Response ~ Background +
Stimulus*I(Intensity - ",i,")*Location")), random = ~ 1|Animal/Cell,
family = neg.bin(theta = 0.8035), data = Rec)
```

Intensity	Contralateral	Ipsilateral	*t*-value	*P*-value
Contralateral vs. Ipsilateral Sound - Congruent Stimulus				
20	0.41	0.37	5.02	P < 0.001
34	0.44	0.38	8.30	P < 0.001

48	0.47	0.40	11.86	P < 0.001
62	0.51	0.42	11.45	P < 0.001
76	0.54	0.44	9.23	P < 0.001
90	0.58	0.46	7.58	P < 0.001
Contralateral vs. Ipsilateral Sound - Incongruent Stimulus				
20	0.36	0.36	0.24	0.808
34	0.40	0.37	2.95	0.003
48	0.45	0.39	6.81	P < 0.001
62	0.50	0.40	8.51	P < 0.001
76	0.56	0.42	8.11	P < 0.001
90	0.62	0.43	7.48	P < 0.001
Congruent vs. Incongruent - Contralateral Sound				
20	0.41	0.36	4.20	P < 0.001
34	0.44	0.40	4.07	P < 0.001
48	0.47	0.45	2.87	0.004
62	0.51	0.50	0.59	0.554
76	0.54	0.56	-0.93	0.353
90	0.58	0.62	-1.69	0.091
Congruent vs. Incongruent - Ipsilateral Sound				
20	0.37	0.36	0.64	0.517
34	0.38	0.37	1.29	0.200
48	0.40	0.39	2.04	0.041
62	0.42	0.40	2.12	0.034
76	0.44	0.42	1.81	0.070
90	0.46	0.43	1.56	0.118

Table 6-2. Results of the generalised linear mixed effects model fitted to spike count data from the categorisation task. Modelled spike counts for congruent and incongruent stimuli are shown according to whether the auditory stimulus component was presented in the contra- or ipsilateral hemifield. The *P*-values and t-values are for the following comparison: contra- vs. ipsilateral sound and congruent vs. incongruent.

6.4.1.3 The Effect of Spatial Congruency on Enhanced and Depressed Responses

A negative binomial generalised linear mixed effects model was fitted to the spike count data obtained from the categorisation task to investigate the effect of stimulus congruency on evoked activity according to whether units exhibited a response enhancement or depression to multisensory stimulation. The model was fitted in R using the code below. Consecutive re-levelling of the model for response modulation and adjustment of the intercept for sound intensity yielded the results in Table 6-3.

```
> mod1 <- glmmPQL(formula(paste("Response ~ Background +
Stimulus*I(Intensity - ",i,")*Location*RM")), random = ~
1|Animal/Cell, family = neg.bin(theta = 0.8058), data = Rec)
```

Intensity	Congruent	Incongruent	*t*-value	*P*-value
Enhanced Units: Congruent vs. Incongruent				
20	0.50	0.45	-2.32	0.020
34	0.53	0.49	-2.17	0.030
48	0.55	0.53	-1.49	0.135
62	0.57	0.57	-0.19	0.850
76	0.60	0.62	0.75	0.453
90	0.62	0.67	1.21	0.223
Depressed Units: Congruent vs. Incongruent				
20	0.36	0.32	-3.46	< 0.001
34	0.40	0.36	-3.42	< 0.001
48	0.43	0.41	-2.38	0.017
62	0.47	0.47	-0.52	0.602
76	0.52	0.54	0.64	0.521
90	0.57	0.61	1.23	0.219

Table 6-3. Results of the generalised linear mixed effects investigating the effects of stimulus congruency on spike counts according to whether units exhibited a response enhancement or depression to multisensory stimulation. Modelled spike counts for congruent and incongruent stimuli are shown for enhanced and depressed units. The *P*- and *t*-values are shown for the following comparison: congruent vs. incongruent.

6.4.1.4 The Effect of Spatial Congruency Spike Timing and Regularity

To investigate the effect of stimulus condition (i.e. congruent vs. incongruent) on spike timing and regularity, separate linear mixed effects models were fitted to the mean spike latency and Fano Factor values. The models were fitted using the R code shown below. Consecutive adjustment of the intercept for sound intensity yielded the results shown in Tables 6-4, 6-5. A Shapiro-Wilk test was performed (mean spike latencies: W = 0.99, P = 0.93; Fano Factor: W = 0.94, P = 0.53) and Q-Q plots (Figs. 6-9, 6-10) were generated to check the model fit, all of which showed that the residuals were normally distributed and, therefore, demonstrated a good fit.

```
> mod.ml <- lme(formula(paste("myValue ~ Stimulus*I(Intensity -
",i,")")), random = ~ 1|Animal, data = Data)
```

Intensity	**Mean Spike Latency (ms)**		***t*-value**	***P*-value**
dB SPL	Congruent	Incongruent		
20	63.6	64.0	0.29	0.77
34	60.9	61.8	0.90	0.37
48	58.2	59.5	1.81	0.07
62	55.5	57.3	2.36	0.02
76	52.8	55.1	2.23	0.03
90	50.0	52.8	1.99	0.05

Table 6-4. Results of the generalised linear model fitted to the mean spike latency data. Modelled latencies for congruent and incongruent stimuli are shown across sound intensities, along with the *P*- and *t*-values for the congruent vs. incongruent comparison.

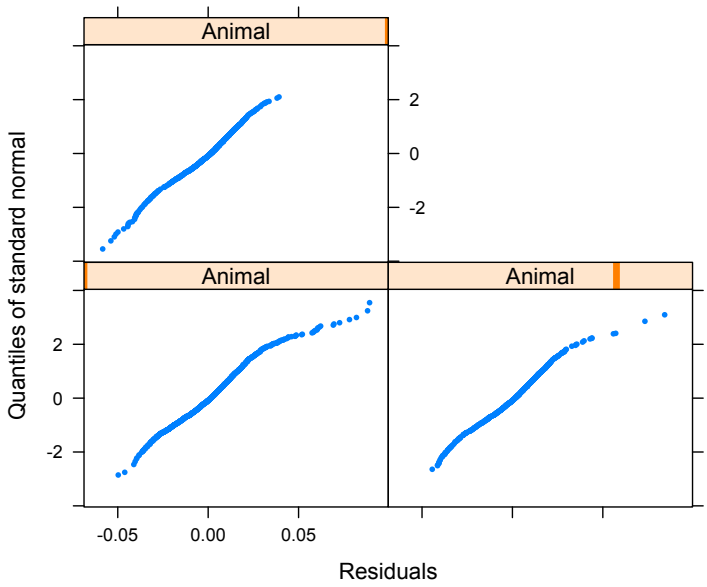


Figure 6-9. Q-Q plot of residuals from the linear mixed effects model fitted to the mean spike latency data. Plots for individual animals are shown in separate panels. The points falling on a straight line indicates that the residuals are normally distributed.

```
> mod.ff <- lme(formula(paste("1/(myValue^(1/5)) ~
Stimulus*I(Intensity - ",i,")")), random = ~ 1|Animal|Cell, data =
Data[Data$myValue>1,])
```

Intensity	Fano Factor		*t*-value	*P*-value
dB SPL	Congruent	Incongruent		
20	2.01	1.93	1.27	0.21
34	2.01	1.96	1.11	0.27
48	2.01	1.99	0.61	0.54
62	2.02	2.02	-0.15	0.88
76	2.02	2.05	-0.65	0.51
90	2.02	2.08	-0.88	0.38

Table 6-5. Results of the generalised linear model fitted to the Fano Factor values. Modelled values for congruent and incongruent stimuli are shown across sound intensities, along with the *P*- and *t*-values for the congruent vs. incongruent comparison.

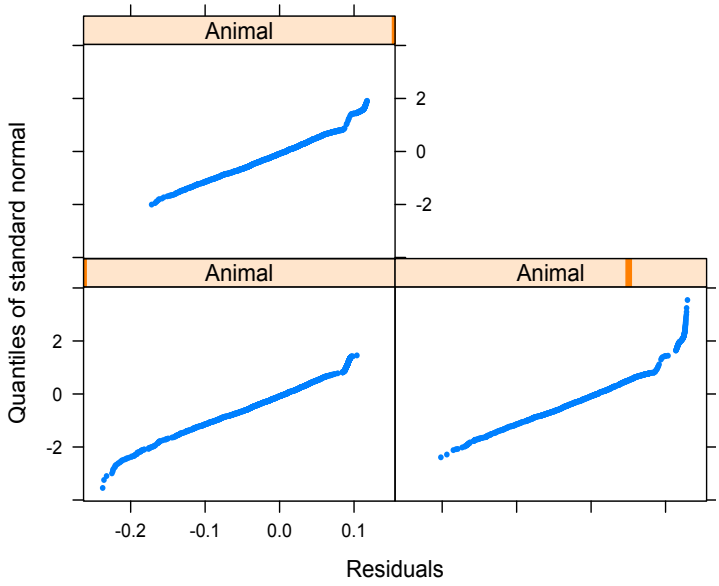


Figure 6-10. Q-Q plot of residuals from the linear mixed effects model fitted to the Fano factor values. Plots for individual animals are shown in separate panels. The points falling on a straight line indicates that the residuals are normally distributed.

6.4.1.5 The Effects of Spontaneous and Evoked Activity on Approach-to-Target Performance

To investigate the effects of spontaneous and evoked activity (i.e. the number of spikes in a 150 ms window before or following stimulus onset, respectively) on approach-to-target response correctness within the different stimulus intensities, conditions (congruent and incongruent) and auditory stimulus component locations relative to each unit (-30° and +30°), a generalised linear mixed effects model for Bernoulli distributed responses was fitted to the data using the R code below. Consecutive re-levelling of the model for stimulus location and condition and adjustment of the intercept for sound intensity yielded the results shown in Table 6-6.

```
modB3.me <- glmmPQL(Correctness ~ Bgd_Activity + Res_Activity +
Loc*Stimilus_Type*Intensity, random = ~ 1|Animal/Cell, family =
binomial, data = Rec)
```

	Evoked Activity			**Stimulus Type**			**Stimulus Location**		
dBSPL	Value	t-value	*p*-value	Value	*t*-value	*p*-value	Value	*t*-value	*p*-value
	Congruent Stimulus, ipsilateral auditory component								
20	0.028	-0.24	0.002	0.43	9.61	<0.001	-0.12	-3.12	0.002
34	0.027	2.17	0.03	1.22	38.9	<0.001	-0.05	-1.99	0.05
48	0.025	1.96	0.05	2.01	65.7	<0.001	0.02	0.57	0.57
62	0.024	1.26	0.21	2.81	64.6	<0.001	0.08	1.86	0.06
76	0.023	0.85	0.40	3.60	58.1	<0.001	0.15	2.33	0.02
90	0.022	0.61	0.54	4.39	53.3	<0.001	0.22	2.55	0.01
	Congruent Stimulus, contralateral auditory component								
20	-0.014	-0.84	0.40	0.58	12.8	<0.001	0.12	3.12	0.002
34	-0.005	-0.42	0.67	1.25	39.9	<0.001	0.05	1.99	0.05
48	0.004	0.36	0.72	1.93	63.8	<0.001	-0.02	-0.57	0.57
62	0.013	0.78	0.43	2.61	61.1	<0.001	-0.08	-1.86	0.06
76	0.022	0.93	0.35	3.28	54.0	<0.001	-0.15	-2.33	0.02
90	0.032	0.99	0.32	3.97	49.0	<0.001	-0.22	-2.55	0.01
	Incongruent Stimulus, ipsilateral auditory component								
20	-0.072	-2.85	0.004	-0.43	-9.61	<0.001	0.02	0.47	0.64
34	-0.057	-3.09	0.002	-1.22	-38.9	<0.001	-0.02	-0.60	0.55
48	-0.043	-2.80	0.005	-2.01	-65.7	<0.001	-0.07	-2.16	0.03
62	-0.028	-1.62	0.106	-2.81	-64.6	<0.001	-0.11	-2.85	0.004
76	-0.014	-0.58	0.564	-3.60	-58.1	<0.001	-0.16	-2.84	0.005
90	0.001	0.03	0.97	-4.39	-53.3	<0.001	0.21	-2.73	0.006
	Incongruent Stimulus, contralateral auditory component								
20	0.036	1.64	0.10	-0.58	-12.8	<0.001	-0.02	-0.47	0.64
34	0.023	1.42	0.16	-1.25	-39.9	<0.001	0.02	0.60	0.55
48	0.009	0.72	0.47	-1.93	-63.8	<0.001	0.07	2.16	0.03

62	-0.004	-0.29	0.77	-2.61	-61.1	<0.001	0.11	2.85	0.004
76	-0.018	-0.88	0.38	-3.28	-54.0	<0.001	0.16	2.84	0.005
90	-0.032	-1.14	0.25	-3.97	-49.0	<0.001	-0.21	2.73	0.006

Table 6-6. Results of the generalised linear model fitted to the approach-to-target response correctness data from the categorisation task. Statistical values for the fixed effects are shown for each stimulus intensity-location-condition combination.

6.4.1.6 Task Comparison

A negative binomial generalised linear mixed effects model was fitted to the spike count data from the localisation and categorisation tasks to investigate whether the evoked activity differs depending on the context of the task. The model was fitted in R using the code below. Consecutive re-levelling of the model for stimulus location and adjustment of the intercept for sound intensity yielded the results in Table 6-7.

```
mod1 <- glmmPQL(response ~ background + Location*Task*I(Intensity -
20), random = ~ 1|Animal|Cell, family = neg.bin(theta = 0.8064),
data = Rec)
```

Intensity	Localisation	Categorisation	*t*-value	*P*-value
Ipsilateral Stimuli: Localisation vs. Categorisation				
20	0.41	0.38	3.34	0.0008
34	0.44	0.40	4.52	<0.0001
48	0.46	0.42	5.41	<0.0001
62	0.48	0.44	4.54	<0.0001
76	0.50	0.46	3.19	0.0014
90	0.53	0.49	2.32	0.0205
Contralateral Stimuli: Localisation vs. Categorisation				

20	0.53	0.41	9.51	<0.0001
34	0.56	0.45	12.2	<0.0001
48	0.60	0.48	13.9	<0.0001
62	0.64	0.52	11.1	<0.0001
76	0.67	0.56	7.31	<0.0001
90	0.72	0.61	4.88	<0.0001

Table 6-7. Results of the generalised linear mixed effects model fitted to spike count data from the categorisation and localisation tasks. Modelled spike counts for both tasks are shown according to whether stimuli were presented from -30° and +30°. The *P*-values and t-values are for the following comparison: localisation vs. categorisation.

7 General Discussion

7.1 Overview

The generation of a unified sensory percept requires the integration of information from different sensory modalities. Contrary to the traditional view that unisensory signals are first processed by dedicated neural pathways, which are assumed to be largely independent and hierarchical, and later on combined and integrated in higher-order association areas, multisensory integration has now been shown to occur even at the level of the core auditory cortex (Brosch *et al.*, 2005; Ghazanfar *et al.*, 2005; Bizley & King, 2008). However, the effect of early multsiensory interactions on the primary role of the auditory cortex in hearing and, ultimately, their behavioural relevance remains unclear. This is because to date most studies recording neural activity from early sensory cortices have been performed either under general anaesthesia or in awake animals that were not required to perform a behavioural task. Therefore, the present experiments were designed to investigate the role of early cross-model interactions by recording the activity of a large number of units in early auditory cortical processing areas, whilst ferrets performed tasks based on auditory-visual stimulus localisation and categorisation behaviours, respectively.

7.1.1 Behavioural Advantages of Multisensory Processing

The integration of information from different sensory modalities is known to confer wide-ranging behavioural benefits (reviewed in Murray & Wallace, 2012). For example, in humans (Alais & Burr, 2004; Odegaard *et al.*, 2015) and cats (Stein *et al.*, 1989a; Gingras *et al.*, 2009) combining auditory and visual stimuli from the same location has been shown to improve localisation accuracy. In order to explore whether ferrets also integrate multisensory cues in a similar fashion, we tested their ability to localise auditory, visual and audio-visual stimuli presented over the frontal hemifield. Our results showed that the provision of spatially-coincident visual cues improved the accuracy of auditory localisation; however, this effect was observed only when the animals had to select which target to approach in order to receive a reward but not when they merely turned their heads towards the target. This finding is supported by previous work from this laboratory, showing that silencing the primary auditory cortical fields in ferrets impairs approach-to-target localisation accuracy when short-duration noise bursts are used as a stimulus, without affecting head orienting behaviour (Nodal *et al.*, 2010, 2012). This suggests that the underlying neural circuits involved in guiding these responses may not be the same.

Race model analysis of the approach-to-target response times indicated that this multisensory facilitation reflects a mechanism of neural integration of the cues available to each sensory modality. Moreover, we found that ferrets, like humans (Alais & Burr, 2004) integrate inputs from different sensory modalities in a statistically optimal manner, whereby each cue is weighted by its respective reliability. In contrast, multisensory facilitation of head orienting reactions times was

shown to be accounted for by probability summation of the unisensory responses. These findings suggest that fundamentally different neural processes are involved in mediating the effects of these two measures of localisation behaviour. Consideration of the findings discussed above confirmed that the ferret does indeed integrate multisensory information in a manner such as to facilitate localisation accuracy, as in the case of humans and cats referred to above, and should provide a suitable model to explore the neural circuits responsible for multisensory spatial processing. These findings also led to the hypothesis that the superior approach-to-target localisation performance, observed with spatially coincident auditory-visual cues, might involve early auditory cortical processing.

7.1.2 Prevalence of Multisensory Interactions in the Early Auditory Cortex

After establishing that the improved localisation performance was a result of multisensory integration, we recorded neural activity whilst ferrets carried out the audio-visual localisation task to investigate the extent and possible functional significance of multisensory convergence in early auditory cortical processing areas. It is becoming increasingly clear that multisensory interactions also occur in the auditory system (see Introduction). The present study, likewise, found evidence for audio-visual interactions in the ferret auditory cortex. In total, at the single cell level, 16% of the units recorded were influenced by visual stimulation. This value is consistent with those reported for neurophysiological recordings from the anaesthetised ferret early auditory cortex (Bizley *et al.*, 2007; Meredith & Allman,

2015) and the awake macaque core auditory cortex (Kayser *et al.*, 2008).

While some units were driven by visual inputs, a greater proportion of units, at the single cell level, had their auditory responses modulated by visual inputs. However, we also found that units classified as unimodal, at the single cell level, displayed subtle cross-modal modulatory effects that were only detectable at the population level. Whilst some units displayed a response enhancement to concurrent visual stimulation, the majority of units exhibited response suppression and, overall, the effect on the population response was suppressive. This is similar to the suppressive effect on the firing rate of units in the auditory cortex of the Macaque monkey, observed by pairing complex naturalistic audio-visual stimuli (Kayser et al., 2008). In comparison to other brain areas such as the SC, our findings indicate that multisensory activity in early auditory cortical processing areas is characterised by relatively low levels of integration, which most commonly takes the form of a suppression of auditory cortical activity. However, multisensory processing in the SC and auditory cortex do appear to share some principles, such as inverse effectiveness (Meredith & Stein, 1983), given that we likewise found that the magnitude of modulation depends on the efficacy of the unisensory stimuli in evoking activity.

7.1.3 Encoding of Sound Source Location by Populations of Cortical Neurons

In addition to facilitating the investigation of the relationship between multisensory interactions at the neural and behavioural levels, this study also bears on the

fundamental question of how auditory space is represented in the cortex. It is well established that an intact auditory cortex is needed for normal spatial hearing (Kavanagh & Kelly, 1987b; Heffner & Heffner, 1990; Nodal *et al.*, 2010). However, to date, a map of auditory space has not been identified in the cortex. Instead, in keeping with our findings, the majority of neurons are broadly tuned to sounds presented in contralateral space, whereas a smaller proportion are tuned to sounds presented in ipsilateral space, thus creating left- and right-tuned channels within and across hemispheres. However, the finding that unilateral cortical lesions produce a sound localisation deficit that is confined to contralateral space (Malhotra & Lomber, 2007) has led to the proposal of an opponent channel population code of space, whereby sound source azimuth is determined by comparing the activity of the left and right tuned channels within the same hemisphere rather than across hemispheres (Stecker *et al.*, 2005). Work published from this laboratory in anaesthetised ferrets (Keating *et al.*, 2015) supports this hypothesis but also shows that behaviourally-relevant interaural level differences can be decoded from a comparison of activity across hemispheres. Likewise, the results from the present experiment also support the proposed existence of a distributed hemispheric code. In particular, our results suggest a direct link between the right/left balance of activity within the auditory cortex and sound localisation behaviour, since we found that animals were more likely to correctly localise stimuli when activity levels in the contralateral cortex increased and activity levels in the ipsilateral cortex decreased. Moreover, a distributed population code could account for how information about space might be combined with other stimulus features (e.g. stimulus intensity) and for the superior spatial acuity observed in the midline in various species, including humans (Makous

& Middlebrooks, 1990; Stecker *et al.*, 2005) and ferrets (King & Parsons, 1999).

7.1.4 Behavioural Relevance of Multisensory Interactions in the Early Auditory Cortex

Overall, our results do not support a direct relationship between neural activity in the auditory cortex and the visual enhancement of sound localisation behaviour. First, behavioural gains were maximal at 20 dB SPL, the lowest sound intensity tested, whilst the greatest modulation of evoked activity occurred at 34 dB SPL, close to the threshold intensity of units for the auditory stimulus. Second, whereas the probability of a correct localisation response increased as evoked contralateral activity levels increased, overall visual co-stimulation resulted in a suppression of the population activity, indicating the absence of a simple relationship between the level of auditory cortical activity and localisation performance. Moreover, surface plots (Fig. 4-12), illustrating the modelled relationship between the evoked activity and probability of a correct localisation response, reveal that the same level of evoked activity for auditory stimuli and auditory-visual stimuli results in substantially different behavioural performance levels for each stimulus type. Although visual stimuli can increase the amount of stimulus-related information conveyed by the temporal firing patterns of auditory cortical neurons, without necessarily altering their firing rates (Bizley & King, 2008; Kayser *et al.*, 2010), it seems likely that the processing of auditory and visual information in other areas, including the visual cortex or higher order association cortices, also contributes to the observed visual enhancement of localisation performance. In support of this hypothesis, it has been proposed that the

parietal cortex might combine spatial information from different sensory modalities, including audition and vision with their associated discordant spatial frames, with that pertaining to head and eye position; thus a common co-ordinated representation of space is generated that is independent of sensory modality (Stricanne *et al.*, 1996; Andersen *et al.*, 1997). Moreover, fMRI studies have shown that in the parietal cortex, auditory and visual signals are integrated according to their respective reliabilities, which is in keeping with the results of behavioural studies showing statistically optimal cue integration, whereas in primary sensory cortices this was not the case (Rohe & Noppeney, 2016).

The question as to the functional significance of multisensory interaction in early auditory cortical processing areas therefore remains. Our results showing that the greatest modulation of evoked activity occurred near to the threshold intensity for the auditory stimulus, together with those suggesting that animals were more likely to make a correct response when the signal-to-noise ratio is high, suggest that one function of visual suppression of auditory cortical activity might be to provide contextual cues that facilitate stimulus detection. In support of this hypothesis, studies have shown that irrelevant visual (Odgaard *et al.*, 2004) stimuli can not only enhance the detection of faint tones but also increase their perceived loudness.

7.1.5 The Use of Spatial Cues in Merging Sensory Stimuli

The integration of sensory inputs from different modalities not only confers behavioural advantages, including enhanced localisation accuracy and reduced

response times but is also essential for other aspects of perception, including scene analysis, by merging different sensory stimuli and resolving conflicts between them. Psychophysical studies show that humans use spatial and temporal cues, as well as semantic information – where appropriate – for this purpose (reviewed by Stein, 2012). In the context of these studies, we designed a second behavioural paradigm that tested the ability of ferrets to categorise temporally coincident auditory-visual stimuli as one object if they were spatially congruent or as discrete objects if they were spatially incongruent. Our results demonstrated that ferrets, like humans, are also capable of using spatial cues to determine whether different stimuli belong to the same object and therefore should be merged or to different objects.

7.1.6 How the Relative Location of Stimuli Affects Multisensory Integration

To explore how the relative location of the stimuli affects multisensory integration in the auditory cortex, we recorded neural activity during the performance of the categorisation task. Electrophysiological recordings from multisensory neurons in the SC have identified spatial coincidence as a key principle by which different sensory inputs interact to govern spiking activity (King & Palmer, 1985; Meredith & Stein, 1986a; Meredith *et al.*, 1987). The SC is characterised by the presence of topographically-aligned maps of auditory and visual space. Consequently, it is relatively straightforward to envisage how different modality signals arising from a particular direction – and therefore potentially from the same source – can be combined by the same neurons. However in the cortex, there are fundamental

differences in the way that space is represented across sensory systems. In the visual system, stimulus location is represented topographically, whereas in the auditory cortex, stimulus location is thought to be encoded by a distributed population code (Stecker et al., 2005). Therefore, it is less clear how auditory and visual spatial representations are reconciled in the auditory cortex. The categorisation task results showed that the presentation of spatially incongruent stimuli had a modulatory effect on the population response, whereby the presence of the spatially incongruent visual stimulus suppressed the auditory response. Although not directly tested, it is feasible that a direct relationship between the magnitude of this modulatory effect and the spatial disparity of stimuli could provide a plausible means for auditory and visual spatial information to be combined in the auditory cortex. In support of this hypothesis, studies of ventriloquism (Zwiers et al., 2003) have illustrated that vision can alter the perceived location of sound sources in order to resolve short-term spatial conflicts between these modalities. Although the neural basis for ventriloquism is not fully understood, event-related potential and fMRI measurements have revealed changes in the balance of activity levels between the left and right hemispheres on trials in which participants experienced a shift in perceived sound location in the direction of a misaligned visual stimulus (Bonath et al., 2007, 2014). Our categorisation task results demonstrated an analogous change, whereby the presence of a spatially discordant visual stimulus suppressed activity in the cortex contralateral to the sound, whereas activity in the ipsilateral cortex was unaffected. Assuming a hemispheric opponent channel code of sound source azimuth, this effect could distort the neural representation of auditory space and thereby alter perception.

7.1.7 Activity in the Early Auditory Cortical Processing Areas is more closely tied to the Physical Properties of the Stimulus rather than the Animal's Perception.

In addition to facilitating the understanding of how inputs from different modalities are combined in the auditory cortex, this study also enabled us to explore several other aspects of early auditory cortical processing such as whether neural activity during task performance is more closely tied to the physical properties of the stimuli or the animal's percept. In support of the latter, neural activity recorded from the auditory cortex, including primary areas, during the performance of various auditory tasks has been shown to be informative about the animal's impending sensory decision (Niwa *et al.*, 2012; Bizley *et al.*, 2013; Tsunada *et al.*, 2015). In contrast to these findings, our results showed that, in the context of a multisensory task, cortical activity was not informative about the animal's behavioural response but was informative about the physical properties of the auditory stimulus. This finding supports the hypothesis that the integration of auditory and visual inputs predominantly occurs in higher-order association cortices.

Finally, we were also able to investigate whether neural activity in the auditory cortex is task-dependent. Electrophysiological recordings from the ferret auditory cortex have shown that when different tasks are performed sequentially (e.g. tone discrimination and detection) (Fritz *et al.*, 2005) or when the task reward structure changes (e.g. approach and avoidance) (David *et al.*, 2012), neurons respond differently to identical tones manifest as a change in their spectro-temporal receptive fields. Similarly, in the present experiment, a comparison of the recordings from the localisation and categorisation tasks showed that the activity evoked by

identical stimuli varied between tasks, whereby the highest activity levels were observed for the localisation task. A population decoding analysis likewise showed that the neural activity was informative about the behavioural context of the task. Together, these findings suggest that neural activity in early auditory cortical processing areas can reflect both the physical properties of stimuli as well as their behavioural context.

7.1.8 Conclusion

In summary, our behavioural results show that the integration of auditory and visual cues results in significantly more accurate and faster localisation responses and that spatial cues can be used to merge different sensory stimuli and resolve conflicts between them. Thus, validating our experimental model as a foundation to explore multisensory processing. Our electrophysiology results support the existence of multisensory interactions in early auditory cortical processing areas in awake behaving animals and suggest that they may subserve specific functions such as stimulus detection and conflict resolution. In contrast, we found no evidence for their involvement in sound localisation. In comparison to other brain regions, our results suggest that early auditory cortical areas are characterised by relatively low levels of multisensory activity and that the primary function of the early auditory cortex is as an auditory feature detector, which can operate in a task-dependent manner.

7.2 Future Work

Although we have focused on the analysis of single unit activity (SUA), multisensory interactions can be investigated at different levels, from the activity of individual neurons to aggregate measures, including local field potentials (LFPs). Therefore, in addition to recording SUA, which comprises the high frequency content of the raw signal (>300 Hz) and reflects the firing of action potentials, we also recorded LFPs, which comprise the low frequency content of the raw signal (<300 Hz) and reflects the membrane potentials of neurons in the vicinity of the recording site. Consequently, LFPs and SUA can together provide an insight into both the synaptic input to and resultant output from a given region (Pesaran, 2009; Sarko *et al.*, 2013). Whilst, relative to the study of unisensory processing, LFP analysis has only recently been applied to the study of multisensory processing, it has already started to provide new mechanistic insights. For example, visual and somatosensory (Lakatos *et al.*, 2007; Schroeder *et al.*, 2008) inputs have been shown to reset the phase of on-going neural oscillations in the auditory cortex and, consequently, amplify or suppress neural responses depending on whether the arrival of auditory inputs is co-incident with periods of high or low excitability, respectively. Furthermore, such phase locking has been linked to improved response times in an auditory-visual selective attention task (Lakatos *et al.*, 2008b). Likewise, future work will also focus on analysing LFP data to investigate how the amplitude, frequency and phase of LFP oscillations in the auditory cortex are influenced by visual stimulation during performance of both tasks. This may provide additional insights into the way visual stimuli interact with activity in the auditory cortex.

The results presented in this thesis suggest that a clear relationship does not exist between multisensory activity, in terms of the modulation of spike rate, timing and regularity, and the visual enhancement of sound localisation ability. Therefore, the question arises as to which other areas could act as a substrate for this effect. Studies of multisensory processing in the ferret auditory cortex highlight the AEG, which has been implicated in the spatial processing pathway, as a potential target. First, it has been shown to contain the highest proportion of units influenced by visual stimulation (ADF and AVF) (Bizley *et al.*, 2007). Second, neural activity in this region has been found to be the most informative, at the level of the auditory cortex, about the azimuthal location of auditory, visual or paired auditory-visual stimuli (Bizley & King, 2008). Third, it has been shown to be innervated by a region of extrastriate visual cortex that has been implicated in spatial processing (Philipp *et al.*, 2006; Bizley *et al.*, 2007). Alternatively, a variety of studies have highlighted the parietal cortex as a further future target. For example, its activity has been shown to predict sound localisation behaviour (Zatorre *et al.*, 2002) and its inactivation, using trans-cranial magnetic stimulation, has been found to introduce systematic errors in spatial hearing (Lewald *et al.*, 2002). The ventral parietal cortex has also been shown to contain multimodal neurons, which have overlapping auditory and visual spatial receptive fields (Schlack *et al.*, 2005) and, intriguingly, to represent stimuli in eye and head reference frames (Stricanne *et al.*, 1996). Moreover, it has been proposed to provide an interface between sensory and motor areas, whereby inputs from different sensory modalities converge with efferent copies of signals from motor areas to code the spatial location of goals for movement (Andersen *et al.*, 1997).

Our analysis suggests that sound source location can be encoded by a

hemispheric model. However, work from our laboratory and others also supports the operation of an intra-hemispheric opponent channel model of sound source localisation (Stecker *et al.*, 2005; Keating *et al.*, 2015). In support of the latter model, we also found contra- as well as ipsilateral tuned neurons within each hemisphere. Therefore, future work will aim to explore the relative contribution of each model to spatial hearing by comparing the ability of a population decoder to classify sound location based on the difference in the mean population activity of neurons in each hemisphere and based on the difference in the mean population activity of neurons that preferred contra and ipsilateral stimuli. Furthermore, because neural activity was recorded from animals during task performance in this study, we will also be able to explore whether the balance of activity in right/left channels within or across hemispheres is more closely tied to the animals localisation behaviour.

Sensory deprivation, particularly in early life, has been shown to result in a phenomenon known as cross-modal plasticity, whereby sensory cortices are reorganised so that inputs from the lost sense are replaced with those from the remaining intact senses, either by the ingrowth of new connections or the unmaking of existing connections (for review see Bavelier & Neville, 2002). In some cases, this has been shown to result in enhanced performance of the remaining intact sensory modalities. For example, studies in humans and cats have shown enhanced detection of peripheral visual stimuli (for review see Bavelier *et al.*, 2006). However, cross-plasticity in sensory cortices has predominantly been studied at the single neuron level using anesthetised animals. Therefore, we propose that new mechanistic insights into cross-modal plasticity in sensory cortices, including the time course and extent of cross-modal re-organisation following a hearing loss and, in addition, its

correlation with task performance and training, could be provided by adapting the localisation paradigm used in the present study. For example, neural activity could be recorded from the auditory cortex of animals with either a permanent (e.g. cochlea lesion) or reversible (e.g. ear plug) hearing loss whilst they localise visual stimuli.

In the present study, we have demonstrated responses in the auditory cortex due to visual stimuli; likewise, auditory inputs have been shown to influence visual cortical activity (Iurilli *et al.*, 2012) and, in addition, a number of studies have shown reciprocal cortico-cortical projections between primary auditory and visual cortices (Bizley *et al.*, 2007; Iurilli *et al.*, 2012). Therefore, in order to gain a more complete picture of auditory-visual interactions in early sensory cortices and their functional relevance, auditory and visual cortices could both be implanted with multi-electrode arrays, which would enable neural activity from both regions to be recorded simultaneously during task performance. Finally, optogenetic manipulation of the visual cortex could be combined with neurophysiological recordings from the auditory cortex, or vice versa, to explore the ensuing effects of manipulating activity patterns in a temporally precise manner during task execution and thereby unravel the contribution of different areas to the behavioural performance and perception.

A APPENDIX

A.1 Organisation of the ferret auditory cortex

The auditory cortex in the ferret is located on the ectosylvian gyrus (EG) of the temporal lobe and is conventionally divided into several fields (Wallace *et al.*, 1997; Nelken *et al.*, 2004; Bizley *et al.*, 2005) (Figs. 2-6; A-1).

The middle ectosylvian gyrus (MEG) comprises two fields: the primary auditory cortex (A1) and the anterior auditory field (AAF), which extends into the ventral wall of the suprasylvian sulcus (sss) (Kowalski *et al.*, 1995; Wallace *et al.*, 1997). Neurons in this region typically exhibit short response latencies and sharp frequency tuning and, in addition, are tonotopically organised (Phillips *et al.*, 1988; Kowalski *et al.*, 1995; Bizley *et al.*, 2005). The tonotopic organisation of the MEG in the ferret has been shown to differ from that found in other mammalian species. In the ferret, parallel frequency gradients have been described in A1 and AAF, whereby high frequencies are located dorsally at the apex of the MEG and low frequencies are located ventrally at the boundary with the posterior EG (PEG) (Bizley *et al.*, 2005). In contrast in other species, including the cat (Reale & Imig, 1980), gerbil (Thomas *et al.*, 1993), guinea pig (Redies *et al.*, 1989) and mouse (Guo *et al.*, 2012), opposing

rostrocaudally oriented frequency gradients have been found in A1 and AAF, which meet to form a high frequency region located at the interface of these fields.

The posterior EG (PEG) comprises two further tonotopically-organised fields (Nelken *et al.*, 2004; Bizley *et al.*, 2005): the posterior suprasylvian field (PSF), which is located caudally, towards the posterior sss, and the posterior pseudosylvian field (PPF), which is located rostrally, towards the pseudosylvian sulcus (pss). The low frequency areas of PSF and PPF are continuous with the low frequency area of A1, thereby making it difficult to distinguish a boundary between the PEG and MEG based on frequency tuning alone. However, neurons in the PSF and PPF have been shown to differ from those in A1 and AAF in a number of other respects. First, they exhibit longer response latencies and a greater proportion of them display sustained responses (Bizley *et al.*, 2005). Second, they demonstrate a higher selectivity for task-relevant stimuli (Atiani *et al.*, 2014). Third, these fields show different best temporal response windows to pitch, timbre and location (Walker et al., 2011). Furthermore, they can be distinguished by cytochrome oxidase activity and neurofilament antibody SMI_{32} staining and, in addition, by their connectivity patterns (Pallas & Sur, 1993; Bajo *et al.*, 2007; Bajo, Nodal, Bizley, *et al.*, 2010b; Bizley *et al.*, 2015). Ventral to PPF and PSF, a further auditory field has been described, ventral posterior (VP), based on its cytoarchitecture and connectivity (Pallas & Sur, 1993; Bajo *et al.*, 2007; Bizley *et al.*, 2015), although to date it has not been characterised functionally.

Two further auditory fields have been described in the anterior EG (AEG): the anterior dorsal field (ADF), which is located close to AAF, and the anterior ventral field (AVF), which is located more ventrally towards the anterior bank.

Unlike fields in the MEG and PEG, both fields in the AEG lack clear tonotopic order and contain neurons that exhibit poor frequency tuning. Furthermore, ADF and AVF are strongly interconnected and contain some neurons that present strong visual responses. In terms of connectivity, ADF is more related to AAF, whereas AVF is more related to VP (Manger *et al.*, 2004; Bizley *et al.*, 2005; Bajo, Nodal, Bizley, *et al.*, 2010b).

Finally, a number of surrounding areas located in the walls of the sulci, including sss and pss, have been defined as multisensory or higher-order sensory/associative areas (Manger *et al.*, 2004, 2005; Keniston *et al.*, 2009; Bizley *et al.*, 2015). Although the different fields have been relatively well defined in the ferret cortex, no systematic representation has been found for sound features such as periodicity, timbre, or spatial location other than frequency. Consequently it has been suggested that the different attributes of sound may be represented by a distributed population code across cortical fields (Nelken *et al.*, 2004; Bizley *et al.*, 2009).

Using the cat as an animal model, Winer's classical studies of thalamo-cortical, cortico-thalamic and cortico-cortical connectivity have divided the auditory cortical fields into four main systems: tonotopic, non-tonotopic, multisensory and limbic (reveiwed in Schreiner & Winer, 2007; Winer & Lee, 2007; Lee & Winer, 2011). The tonotopic system in the cat comprises A1, AAF and three posterior fields (P, PVP and Ve), which are all tonotopically organised. This system is the main recipient of the lemniscal thalamic and cortical projections and can be considered as an early auditory processing system in the cortex. Following the same criteria, primates have been shown to have three cortical fields that are tonotopically organised in the same antero-posterior axes and that share the same frequency

borders so are, therefore, also considered to be core areas (Kaas & Hackett, 2000; de la Mothe *et al.*, 2006a, 2006b). According to Winer's criteria, the electrophysiological properties (Bizley *et al.*, 2005), anatomical connectivity (Bajo *et al.*, 2007; Bajo, Nodal, Bizley, *et al.*, 2010b; Bizley *et al.*, 2015), and tonotopic organisation of AAF, PSF and PPF, strongly suggests that these fields in the ferret auditory cortex form the tonotopic or early auditory processing system.

A.2 Location of Electrode Arrays

The purpose of this study was to explore the extent and functional significance of multisensory integration in early auditory processing cortices. Therefore, we aimed to implant electrode arrays over the MEG and/or the dorsal PEG. This region was chosen for implantation because it typically contains neurons that both exhibit characteristics associated with early cortical processing areas, including short auditory response latencies and frequency sensitivity (Kelly & Judge, 1994; Wallace *et al.*, 1997; Bizley *et al.*, 2005), and also display some sensitivity to visual stimulation (Bizley *et al.*, 2007; Bizley & King, 2009). The location for implantation was determined intra-operatively from the sulcal pattern on the cortical surface. Post mortem examination of the surface imprints of the arrays confirmed the correct placement in all cases (e.g. Figs. 2-6 and A-1). Following sufficient data collection at each depth, electrodes in the array were advanced to sample different units.

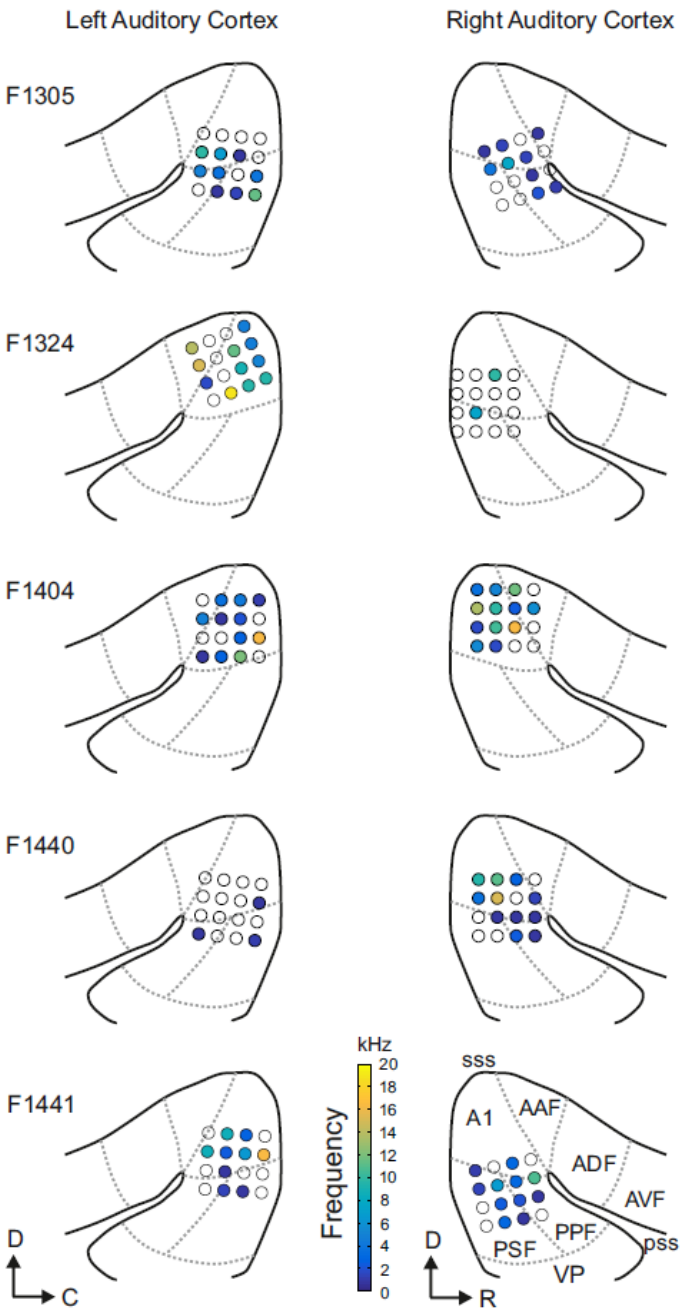


Figure A-1. Schematics showing the location of electrode arrays implanted on the surface of the EG for all animals superimposed over a standard EG silhouette. In some cases, the actual contour of the sulci and array had to be rotated to match the standard EG silhouette. The black contours represent sss and pss and the dotted lines represent the different cortical fields (A1, primary cortex; AAF, anterior field; ADF, anterior dorsal field; AVF, anterior ventral field; PSF, posterior suprasylvian field, PPF posterior pseudosylvian field, VP, ventroposterior field. Electrodes are represented by the circles (empty circles represent electrodes from which no activity was recorded). Electrodes sites are colour coded according to their average best frequency. Abbreviations: C, caudal or posterior; D, dorsal; and R, rostral or anterior.

Three arrays (both sides in the case of ferret 1404 and the left side in the case of 1324) were located over the MEG, whereas seven arrays were located over the MEG and dorsal PEG boundary, making it difficult to assign each recording site to a particular cortical field because of individual variations in gyrus morphology, frequency tuning and the relative angle between the direction of the electrode tract and the cortical surface. Furthermore, as already mentioned, the low frequency areas of PSF and PPF are continuous with the low frequency area of A1, thereby making it difficult to distinguish a boundary between the PEG and MEG based on frequency tuning alone. Moreover, due to the low density of our recordings, we found no clear tonotopic gradients across the fields over which arrays were placed (Figs. A-1 and A-2). Consequently, for the aforementioned reasons, we have not attempted to assign recording sites to a particular cortical field since this would not be possible without using specific neuromarkers (e.g. neurofilament SMI_{32}) (Bajo *et al.*, 2007) or performing a detailed investigation of the underlying connectivity of these regions. Unfortunately, these anatomical methods are not compatible with chronic electrophysiological implants.

The lack of clear tonotopy is illustrated by Figure 4-2, which shows the location and average frequency selectivity of units recorded in the left auditory cortex of animal F1324. Assuming a high to low tonotopic gradient running along a dorsal-ventral axis, the most ventral recording sites (see electrodes 10, 8 & 4) might be expected to present low best frequencies. However, their values ranged from 8.9 to 22 kHz. Furthermore, in comparison, some of the more dorsal recording sites displayed lower best frequencies values (e.g. electrodes 13 (0.6 kHz), 16 (1.4 kHz) and 2 (2.7 kHz)). One possible explanation for this lack of clear tonotopy is that

arrays were not implanted orthogonally relative to the cortical surface, which is noticeable in the histological sections (Fig. A-3) and the 3-dimensional reconstruction (Fig. A-4). Hence when the electrodes were advanced through the cortex it is possible that their tips also moved across the tonotopic axis either within a cortical field or across them. Consequently, in addition to using specific neuromarkers, the accurate assignment of each recording site to a particular field would also require the analysis of the histological sections (Fig. A-3), rather than relying solely on the photographs of the electrode imprints on the cortical surface.

In the ferret, the lateral location of the EG, together with the dimensions of the array, precluded the orthogonal implantation of electrode arrays since this would have resulted in their lateral protrusion, which we avoided for a number of reasons. First, it would have added instability to the implant, particularly when the animals were tethered for the neural recordings. Second, it would have resulted in the animal knocking the implant against the wall of the testing arena when approaching the response locations. Third, it would have disrupted the normal anatomical position of the ears, thereby distorting the external auditory canal and altering the binaural and spectral cues required for sound localisation. Fourth, it would not necessarily have guaranteed the orthogonal entry of all electrodes, since the surface of the EG is curved with a thick cortex at its centre that gradually thins toward the sulci. Finally, orthogonal implantation would have reduced the potential number of recordable units. Assuming that electrodes were inserted at an angle of ~ 60° to the cortical surface and that the cortex is ~ 1500 μm thick, electrodes would have needed to be advanced a distance of 1732 μm to pass through all cortical layers, which equates to a 15% increase in tract length, thereby increasing the number of recordable positions

along each tract, given that we typically advanced electrodes in 100 μm steps. However, this angular displacement would have also resulted in a dorso-ventral shift of ~ 866 μm from the point of cortical entry (see Fig. A-3, right side), which would likely be associated with a change in best frequency.

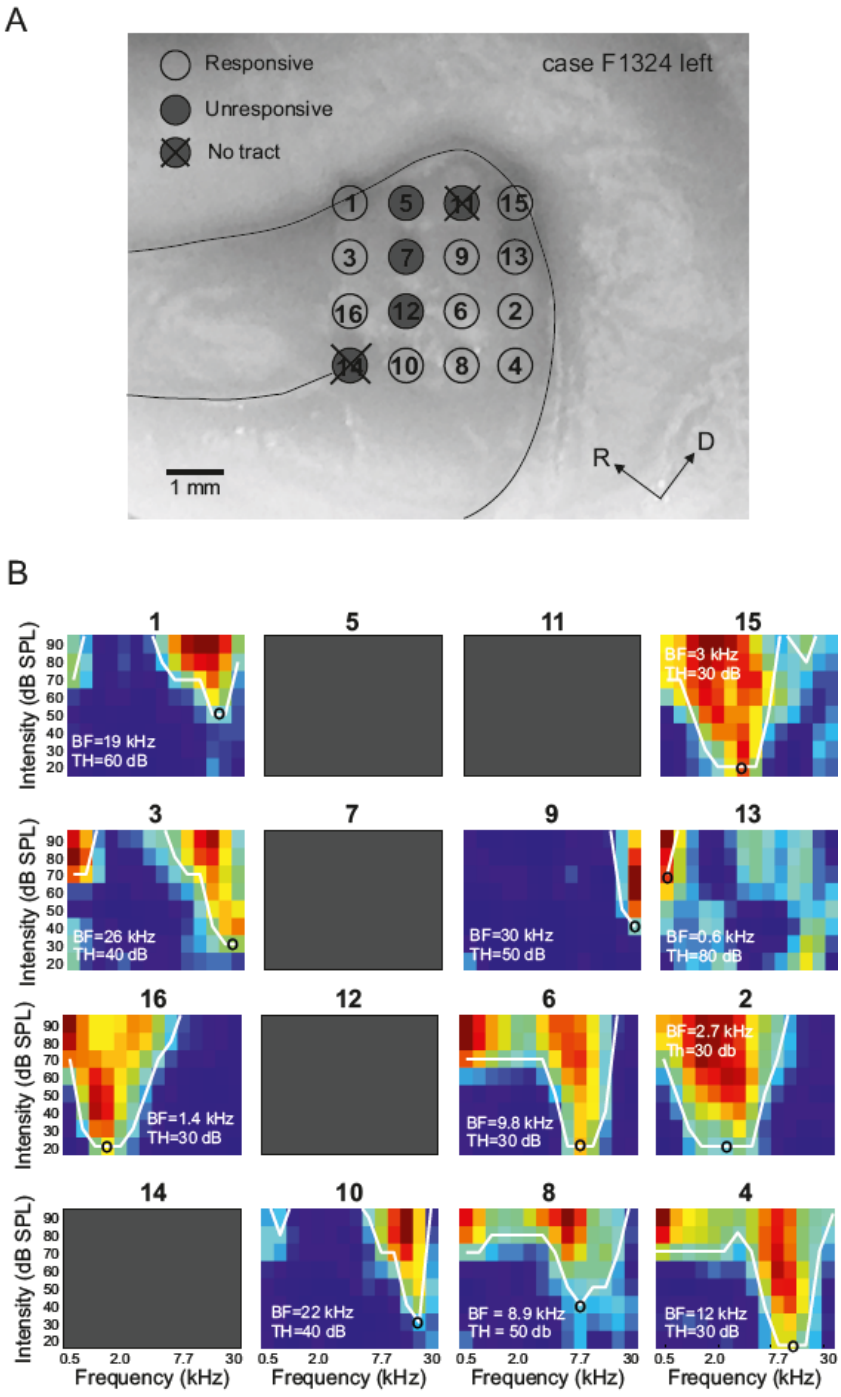


Figure A-2. Location and frequency sensitivity of units recorded from an electrode array implanted over the left cortex (case F1324). (A) Photograph of the left EG surface showing the location of electrode imprints and neuron responsiveness. Numbers represent the mapping of the electrodes onto the recording channels. (B) Frequency response areas (FRAs) of exemplar units from different electrodes. The colour map represents the relative number of spikes evoked for each frequency-sound level combination and the white line delineates the tuning curve (see Section 2.6.4). Best frequency (BF, empty black circle) and threshold intensity (TH) are shown for each unit.

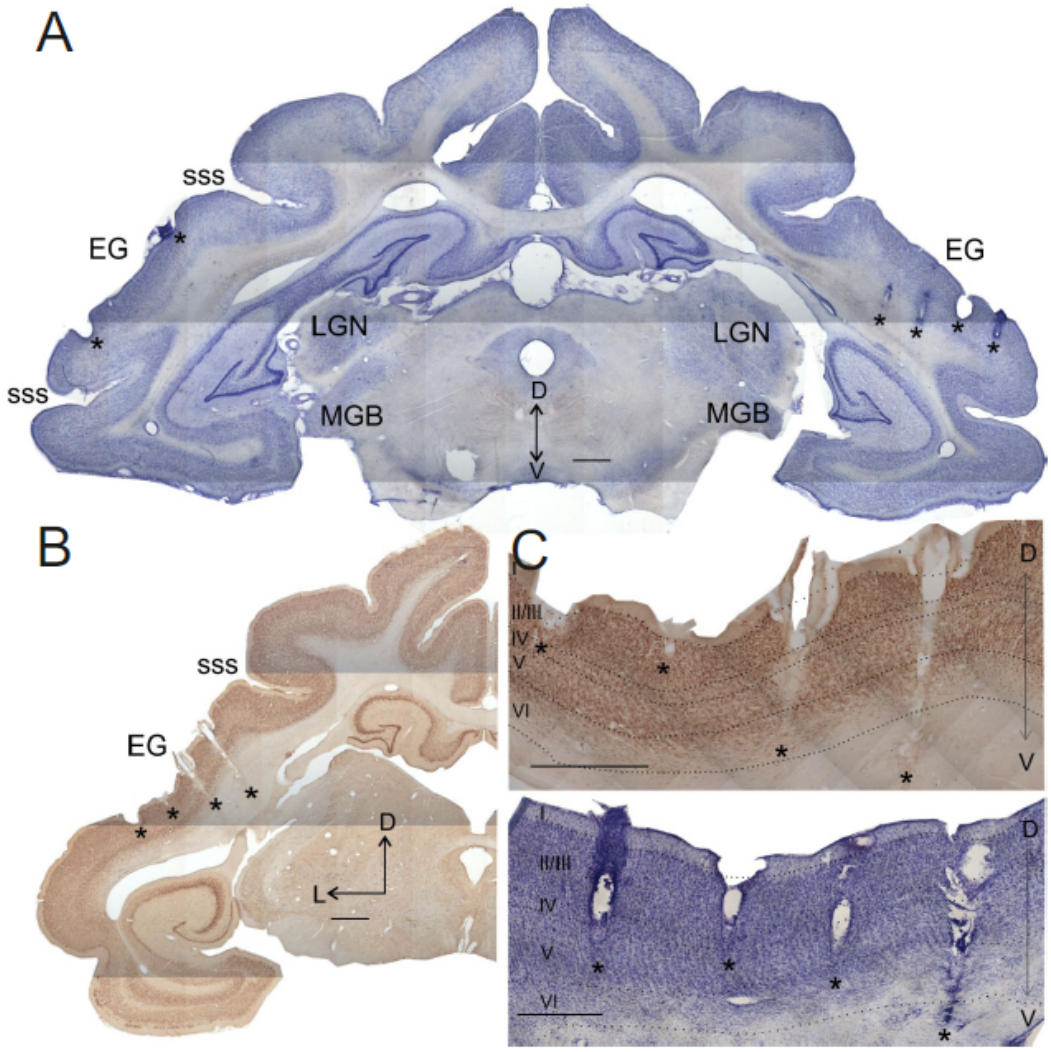


Figure A-3. Photomicrographs of histological coronal sections (case F1440) showing the electrode tracts across the EG bilaterally. (A) Panoramic photomicrograph of a Nissl stained section showing the electrode tracts on the left and right side of the EG. (B) Photomicrograph of a more anterior NeuN stained coronal section from the left side of the same animal. Four electrode tracts can be observed with the tip of the most dorsal track located in the white matter. (C) Higher magnification photomicrographs of the electrode track depicted in A (right side) and B (left side) onto which the cortical laminar boundaries are drawn. Abbreviations: EG, ectosylvian gyrus; D, dorsal; LGN, lateral geniculate body; MGB, medial geniculate body; sss, suprasylvian gyrus, V, ventral; I-VI, cortical layers 1-6. Calibration bars = 1 mm.

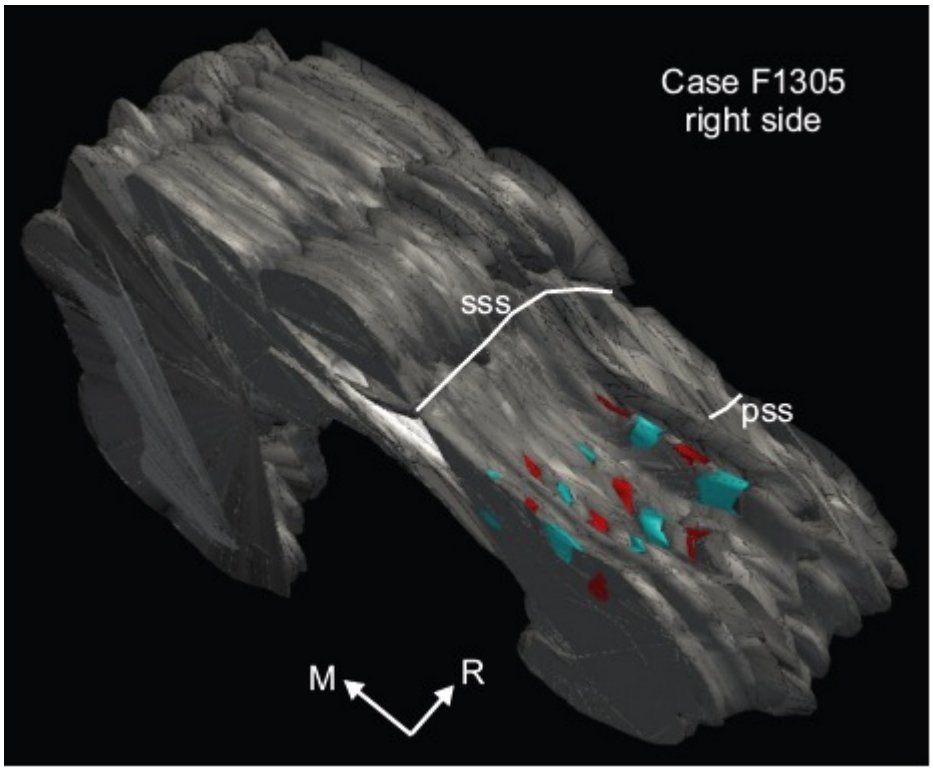


Figure A-4. 3-D reconstruction of the cortex at the level of the EG made from serial histological sections (case F1305). Different electrode tracts are colour coded for better visualisation and white lines are superimposed to highlight sulcal landmarks. Abbreviations: M, medial; pss, pseudosylvian sulcus, R, rostral; sss; suprasylvian sulcus.

A.3 Laminar Location of the Recordings

3-dimensional reconstructions and histological sections showed good correlation with experimental diaries in terms of the electrode locations that yielded responses. At most positions where channels lacked driven responses, we found that this was typically because the electrode had not entered the cortex or it had exited the cortex and entered the white matter. In general, there was agreement between the actual

length of the electrode tracts and their estimated length from summing the distance that each electrode was advanced at the end of every recording cycle. The number of times electrodes were advanced differed for each animal and ranged from 7 to 14 times (F1305: 7; F1440: 9; F1441: 8; F1444: 11 and F1324: 14). This difference was, in part, due to individual variation in the effectiveness of the electrodes penetrating the dura mater. For example, some electrodes appeared to pierce the dura mater more readily, as evidenced by the disruption they caused to layer 1 (Fig. A-3C, bottom panel, left most tract), than others that appeared to have compressed rather than pierced the cortical surface (Fig. A-3C, top panel, right most tract). For those electrodes that the dura mater offered greater resistance, we assumed that the initial position that yielded responses could have been 200-300 μm from the surface.

In general, the cortical layers were well preserved despite electrode penetration. Furthermore, macroscopic inspection of the cortical surface showed that the chronic recordings caused minimal damage (e.g. see Fig. 2-A). Likewise, microscopic inspection of the histological sections confirmed a healthy appearance of the neural tissue except at the cortical surface (e.g. see Fig. A-3C). Moreover, the quality of the neural recordings, together with our behavioural results showing that animals were able to perform the task following implantation, support these findings given that a lesion might be expected to have resulted in an behavioural deficit.

The only reliable indicator that an electrode was in situ in the cortex was the recording of neural activity from it. However, because of the aforementioned factors, we cannot assume that the first location at which activity was recorded was in the most superficial layers. In fact, the assignment of recording locations along each electrode tract by working backwards from the deepest location indicates that most

units were recorded from intermediate and deep layers rather than superficial layers. For this reason and because of the morphological changes that can be introduced by fixation and histological processing, we have decided not to assign each unit to an exact laminar location. We have, however, used the histological reconstructions to ensure that all the recordings were obtained from the auditory cortex and not from deeper structures like the claustrum or hippocampus (Fig A-3), which are situated underneath the auditory cortex.

REFERENCES

Alais, D. & Burr, D. (2004) The ventriloquist effect results from near-optimal bimodal integration. *Curr. Biol.*, **14**, 257–262.

Allen, A.E., Procyk, C.A., Brown, T.M., & Lucas, R.J. (2016) Convergence of visual and whisker responses in the primary somatosensory thalamus (ventral posterior medial region) of the mouse. *J. Physiol.*, **58**, 7250–7257.

Allman, B.L., Keniston, L.P., & Meredith, M.A. (2009) Not just for bimodal neurons anymore: the contribution of unimodal neurons to cortical multisensory processing. *Brain Topogr.*, **21**, 157–167.

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. *Annu. Rev. Neurosci.*, **20**, 303–330.

Angelaki, D.E., Gu, Y., & Deangelis, G.C. (2011) Visual and vestibular cue integration for heading perception in extrastriate visual cortex. *J. Physiol.*, **589**, 825–833.

Atiani, S., David, S. V, Elgueda, D., Locastro, M., Radtke-Schuller, S., Shamma, S.A., & Fritz, J.B. (2014) Emergent selectivity for task-relevant stimuli in

higher-order auditory cortex. *Neuron*, **82**, 486–499.

Bajo, V.M., Nodal, F.R., Bizley, J.K., & King, A.J. (2010a) The non-lemniscal auditory cortex in ferrets: convergence of corticotectal inputs in the superior colliculus. *Front. Neuroanat.*, **4**, 18.

Bajo, V.M., Nodal, F.R., Bizley, J.K., & King, A.J. (2010b) The non-lemniscal auditory cortex in ferrets: convergence of corticotectal inputs in the superior colliculus. *Front. Neuroanat.*, **4**, 18.

Bajo, V.M., Nodal, F.R., Bizley, J.K., Moore, D.R., & King, A.J. (2007) The ferret auditory cortex: descending projections to the inferior colliculus. *Cereb. Cortex*, **17**, 475–491.

Bajo, V.M., Nodal, F.R., Moore, D.R., & King, A.J. (2010) The descending corticocollicular pathway mediates learning-induced auditory plasticity. *Nat. Neurosci.*, **13**, 253–260.

Battaglia, P.W., Jacobs, R.A., & Aslin, R.N. (2003) Bayesian integration of visual and auditory signals for spatial localization. *J. Opt. Soc. Am. A. Opt. Image Sci. Vis.*, **20**, 1391–1397.

Bavelier, D., Dye, M.W.G., & Hauser, P.C. (2006) Do deaf individuals see better? *Trends Cogn. Sci.*, **10**, 512–518.

Bavelier, D. & Neville, H.J. (2002) Cross-modal plasticity: where and how? *Nat. Rev. Neurosci.*, **3**, 443–452.

Bell, A.H. & Corneil, B.D. (2003) Engagement of visual ® xation suppresses sensory responsiveness and multisensory integration in the primate superior

colliculus **18**.

Bell, A.H., Meredith, M.A., Van Opstal, a J., & Munoz, D.P. (2005) Crossmodal integration in the primate superior colliculus underlying the preparation and initiation of saccadic eye movements. *J. Neurophysiol.*, **93**, 3659–3673.

Bertini, C., Leo, F., Avenanti, A., & Làdavas, E. (2010) Independent mechanisms for ventriloquism and multisensory integration as revealed by theta-burst stimulation. *Eur. J. Neurosci.*, **31**, 1791–1799.

Bishop, K.N., McClung, J.R., Goldberg, S.J., & Shall, M.S. (1985) Anatomic and physiological characteristics of the ferret lateral rectus muscle and abducens nucleus. *J. Appl. Physiol.*, **103**, 1706–1714.

Bizley, J.K., Bajo, V.M., Nodal, F.R., & King, A.J. (2015) Cortico-Cortical Connectivity Within Ferret Auditory Cortex. *J. Comp. Neurol.*, **523**, 2187–2210.

Bizley, J.K. & King, A.J. (2008) Visual–auditory spatial processing in auditory cortical neurons. *Brain Res.*, **1242**, 24–36.

Bizley, J.K. & King, A.J. (2009) Visual influences on ferret auditory cortex. *Hear. Res.*, **258**, 55–63.

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. *Cereb. Cortex*, **17**, 2172–2189.

Bizley, J.K., Nodal, F.R., Nelken, I., & King, A.J. (2005) Functional organization of ferret auditory cortex. *Cereb. Cortex*, **15**, 1637–1653.

Bizley, J.K., Walker, K.M.M., Nodal, F.R., King, A.J., & Schnupp, J.W.H. (2013) Auditory cortex represents both pitch judgments and the corresponding acoustic cues. *Curr. Biol.*, **23**, 620–625.

Bizley, J.K., Walker, K.M.M., Silverman, B.W., King, A.J., & Schnupp, J.W.H. (2009) Interdependent encoding of pitch, timbre, and spatial location in auditory cortex. *J. Neurosci.*, **29**, 2064–2075.

Boehnke, S.E. & Phillips, D.P. (2005) The relation between auditory temporal interval processing and sequential stream segregation examined with stimulus laterality differences. *Percept. Psychophys.*, **67**, 1088–1101.

Bonath, B., Noesselt, T., Krauel, K., Tyll, S., Tempelmann, C., & Hillyard, S.A. (2014) Audio-visual synchrony modulates the ventriloquist illusion and its neural/spatial representation in the auditory cortex. *Neuroimage*, **98**, 425–434.

Bonath, B., Noesselt, T., Martinez, A., Mishra, J., Schwiecker, K., Heinze, H.J., & Hillyard, S.A. (2007) Neural Basis of the Ventriloquist Illusion. *Curr. Biol.*, **17**, 1697–1703.

Britten, K., Newsome, W., Shadlen, M.N., Celebrini, S., & Movshon, J. (1996) A relationship between behavoiral choice and the visual responses of neurons in macaque MT. *Vis. Neurosci.*, **13**, 87–100.

Britten, K.H., Shadlen, M.N., Newsome, W.T., & Movshon, J. a (1992) The analysis of visual motion: a comparison of neuronal and psychophysical performance. *J. Neurosci.*, **12**, 4745–4765.

Brosch, M., Selezneva, E., & Scheich, H. (2005) Nonauditory events of a behavioral

procedure activate auditory cortex of highly trained monkeys. *J. Neurosci.*, **25**, 6797–6806.

Brown, C.H. & May, B.J. (2005) Comparative Mammalian Sound Localization. In Popper, A.N. & Fay, R.R. (eds), *Sound Source Localization*. Springer New York, New York, NY, pp. 124–178.

Brugge, J., Reale, R., & Hind, J. (1996) The structure of spatial receptive fields of neurons in primary auditory cortex of the cat. *J. Neurosci.*, **16**, 4420–4437.

Brugge, J.F., Reale, R.A., Jenison, R.L., & Schnupp, J. (2001) Auditory cortical spatial receptive fields. *Audiol. Neurootol.*, **6**, 173–177.

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**, 1065–1083.

Burnett, L.R., Stein, B.E., Chaponis, D., & Wallace, M.T. (2004) Superior colliculus lesions preferentially disrupt multisensory orientation. *Neuroscience*, **124**, 535–547.

Bushara, K.O., Hanakawa, T., Immisch, I., Toma, K., Kansaku, K., & Hallett, M. (2003) Neural correlates of cross-modal binding. *Nat. Neurosci.*, **6**, 190–195.

Calvert, G.A., Bullmore, E.T., Brammer, M.J., Campbell, R., Williams, S.C., McGuire, P.K., Woodruff, P.W., Iversen, S.D., & David, A.S. (1997) Activation of auditory cortex during silent lipreading. *Science (80-.).*, **276**, 593–596.

Campbell, R.A.A., Schulz, A.L., King, A.J., & Schnupp, J.W.H. (2010) Brief sounds evoke prolonged responses in anesthetized ferret auditory cortex. *J.*

Neurophysiol., **103**, 2783–2793.

Campi, K.L., Bales, K.L., Grunewald, R., & Krubitzer, L. (2010) Connections of auditory and visual cortex in the prairie vole (Microtus ochrogaster): evidence for multisensory processing in primary sensory areas. *Cereb. Cortex*, **20**, 89–108.

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

Cappe, C., Rouiller, E.M., & Barone, P. (2012) Cortical and Thalamic Pathways for Multisensory and Sensorimotor Interplay. *Neural Bases Multisensory Process.*, 13–28.

Champoux, F., Tremblay, C., Mercier, C., Lassonde, M., Lepore, F., Gagné, J.-P., & Théoret, H. (2006) A role for the inferior colliculus in multisensory speech integration. *Neuroreport*, **17**, 1607–1610.

Chandrasekaran, C., Lemus, L., & Ghazanfar, A.A. (2013) Dynamic faces speed up the onset of auditory cortical spiking responses during vocal detection. *Proc. Natl. Acad. Sci. U. S. A.*, **110**, E4668-77.

Chandrasekaran, C., Trubanova, A., Stillittano, S., Caplier, A., & Ghazanfar, A.A. (2009) The natural statistics of audiovisual speech. *PLoS Comput. Biol.*, **5**.

Clemo, H.R., Keniston, L.P., & Meredith, M.A. (2012) *Structural Basis of Multisensory Processing: Convergence*, The Neural Bases of Multisensory Processes.

Colavita, F.B. (1974) Human sensory dominance. *Percept. {&} Psychophys.*, **16**, 409–412.

Collett, D. (2002) *Modelling Binary Data*, Second. edn. Chapman & Hall/CRC.

Colonius, H. & Arndt, P. (2001) A two-stage model for visual-auditory interaction in saccadic latencies. *Percept. Psychophys.*, **63**, 126–147.

Cooper, M.H. & Young, P.A. (1976) Cortical projections to the inferior colliculus of the cat. *Exp. Neurol.*, **51**, 488–502.

Corneil, B.D., Van Wanrooij, M., Munoz, D.P., & Van Opstal, A.J. (2002) Auditory-visual interactions subserving goal-directed saccades in a complex scene. *J. Neurophysiol.*, **88**, 438–454.

David, S. V, Fritz, J.B., & Shamma, S.A. (2012) Task reward structure shapes rapid receptive field plasticity in auditory cortex. *Proc. Natl. Acad. Sci. U. S. A.*, **109**, 2144–2149.

Day, M.L. & Delgutte, B. (2013) Decoding Sound Source Location and Separation Using Neural Population Activity Patterns. *J. Neurosci. Off. J. Soc. Neurosci.*, **33**, 15837–15847.

de la Mothe, L.A., Blumell, S., Kajikawa, Y., & Hackett, T.A. (2006a) Thalamic connections of the auditory cortex in marmoset monkeys: Core and medial belt regions. *J. Comp. Neurol.*, **496**, 72–96.

de la Mothe, L.A., Blumell, S., Kajikawa, Y., & Hackett, T.A. (2006b) Cortical connections of the auditory cortex in marmoset monkeys: core and medial belt regions. *J. Comp. Neurol.*, **496**, 27–71.

DeValois, R.L. & DeValois, K.K. (1991) *Spatial Vision*. Oxford University Press.

Doubell, T.P., Baron, J., Skaliora, I., & King, A.J. (2000) Topographical projection from the superior colliculus to the nucleus of the brachium of the inferior colliculus in the ferret: Convergence of visual and auditory information. *Eur. J. Neurosci.*, **12**, 4290–4308.

Eden, U.T. & Kramer, M.A. (2010) Drawing inferences from Fano factor calculations. *J. Neurosci. Methods*, **190**, 149–152.

Eliades, S.J. & Wang, X. (2008) Chronic multi-electrode neural recording in free-roaming monkeys. *J. Neurosci. Methods*, **172**, 201–214.

Eramudugolla, R., McAnally, K.I., Martin, R.L., Irvine, D.R.F., & Mattingley, J.B. (2008) The role of spatial location in auditory search. *Hear. Res.*, **238**, 139–146.

Ernst, M.O. & Banks, M.S. (2002) Humans integrate visual and haptic information in a statistically optimal fashion. *Nature*, **415**, 429–433.

Ernst, M.O., Bresciani, J.-P., Drewing, K., & Bülthoff, H.H. (2004) Integration of Sensory Information Within Touch and Across Modalities. *Proc. 2004 IEEE/RSJ Int. Conf. Intell. Robot. Syst. (IROS 2004)*, 1–8.

Falchier, A., Clavagnier, S., Barone, P., & Kennedy, H. (2002) Anatomical evidence of multimodal integration in primate striate cortex. *J. Neurosci.*, **22**, 5749–5759.

Falchier, A., Schroeder, C.E., Hackett, T.A., Lakatos, P., Nascimento-Silva, S., Ulbert, I., Karmos, G., & Smiley, J.F. (2010) Projection from visual areas V2 and prostriata to caudal auditory cortex in the monkey. *Cereb. Cortex*, **20**, 1529–1538.

Fano, U. (1947) Ionization Yield of Radiations. II. The Fluctuations of the Number of Ions. *Phys. Rev.*, **72**, 26–29.

Felsen, G. & Mainen, Z.F. (2008) Neural Substrates of Sensory-Guided Locomotor Decisions in the Rat Superior Colliculus. *Neuron*, **60**, 137–148.

Fitts, P.M. (1954) The Information Capacity of the Human Motor. *J. Exp. Biol.*, **47**, 381–391.

Fontanini, A. & Katz, D.B. (2006) State-dependent modulation of time-varying gustatory responses. *J. Neurophysiol.*, **96**, 3183–3193.

Freedman, E.G. & Sparks, D.L. (1997) Activity of cells in the deeper layers of the superior colliculus of the rhesus monkey: evidence for a gaze displacement command. *J. Neurophysiol.*, **78**, 1669–1690.

Frens, M.A. & Van Opstal, A.J. (1998) Visual-auditory interactions modulate saccade-related activity in monkey superior colliculus. *Brain Res. Bull.*, **46**, 211–224.

Frens, M.A., Van Opstal, A.J., & Van der Willigen, R.F. (1995) Spatial and temporal factors determine auditory-visual interactions in human saccadic eye movements. *Percept. Psychophys.*, **57**, 802–816.

Fritz, J.B., Elhilali, M., & Shamma, S.A. (2005) Differential dynamic plasticity of A1 receptive fields during multiple spectral tasks. *J. Neurosci.*, **25**, 7623–7635.

Fu, K.-M.G., Shah, A.S., O'Connell, M.N., McGinnis, T., Eckholdt, H., Lakatos, P., Smiley, J., & Schroeder, C.E. (2004) Timing and laminar profile of eye-position effects on auditory responses in primate auditory cortex. *J. Neurophysiol.*, **92**,

3522–3531.

Fu, K.M., Johnston, T.A., Shah, A.S., Arnold, L., Smiley, J., Hackett, T.A., Garraghty, P.E., & Schroeder, C.E. (2003) Auditory cortical neurons respond to somatosensory stimulation. *J. Neurosci.*, **23**, 7510–7515.

Furukawa, S. & Middlebrooks, J.C. (2002) Cortical representation of auditory space: information-bearing features of spike patterns. *J. Neurophysiol.*, **87**, 1749–1762.

Gandhi, N.J. & Katnani, H.A. (2011) Motor functions of the superior colliculus. *Annu. Rev. Neurosci.*, **34**, 205–231.

Gao, P.P., Zhang, J.W., Fan, S.-J., Sanes, D.H., & Wu, E.X. (2015) Auditory midbrain processing is differentially modulated by auditory and visual cortices: An auditory fMRI study. *Neuroimage*, **123**, 22–32.

Garipis, N. & Hoffmann, K.P. (2003) Visual field defects in albino ferrets (Mustela putorius furo). *Vision Res.*, **43**, 793–800.

Ghazanfar, A.A. (2009) The multisensory roles for auditory cortex in primate vocal communication. *Hear. Res.*, **258**, 113–120.

Ghazanfar, A.A., Chandrasekaran, C., & Logothetis, N.K. (2008) Interactions between the superior temporal sulcus and auditory cortex mediate dynamic face/voice integration in rhesus monkeys. *J. Neurosci.*, **28**, 4457–4469.

Ghazanfar, A.A. & Lemus, L. (2010) Multisensory Integration: Vision Boosts Information through Suppression in Auditory Cortex. *Curr. Biol.*, **20**, R22–R23.

Ghazanfar, A.A., Maier, J.X., Hoffman, K.L., & Logothetis, N.K. (2005) Multisensory integration of dynamic faces and voices in rhesus monkey

auditory cortex. *J. Neurosci.*, **25**, 5004–5012.

Ghazanfar, A.A. & Schroeder, C.E. (2006) Is neocortex essentially multisensory? *Trends Cogn. Sci.*, **10**, 278–285.

Gillmeister, H. & Eimer, M. (2007) Tactile enhancement of auditory detection and perceived loudness. *Brain Res.*, **1160**, 58–68.

Gingras, G., Rowland, B. a, & Stein, B.E. (2009) The differing impact of multisensory and unisensory integration on behavior. *J. Neurosci.*, **29**, 4897–4902.

Gold, J.R., Nodal, F.R., Peters, F., King, A.J., & Bajo, V.M. (2015) Auditory Gap-in-Noise Detection Behavior in Ferrets and Humans. *Behav. Neurosci.*,.

Goldring, J.E., Dorris, M.C., Corneil, B.D., Ballantyne, P.A., & Munoz, D.P. (1996) Combined eye-head gaze shifts to visual and auditory targets in humans. *Exp. brain Res.*, **111**, 68–78.

Goodman, D.F.M., Benichoux, V., & Brette, R. (2013) Decoding neural responses to temporal cues for sound localization. *Elife*, **2013**, 1–21.

Grant, K.W. & Seitz, P.F. (2000) The use of visible speech cues for improving auditory detection of spoken sentences. *J. Acoust. Soc. Am.*, **108**, 1197–1208.

Groh, J.M., Trause, A.S., Underhill, A.M., Clark, K.R., & Inati, S. (2001) Eye position influences auditory responses in primate inferior colliculus. *Neuron*, **29**, 509–518.

Guo, W., Chambers, A.R., Darrow, K.N., Hancock, K.E., Shinn-Cunningham, B.G., & Polley, D.B. (2012) Robustness of cortical topography across fields, laminae,

anesthetic states, and neurophysiological signal types. *J. Neurosci.*, **32**, 9159–9172.

Haider, B. & McCormick, D.A. (2009) Rapid Neocortical Dynamics: Cellular and Network Mechanisms. *Neuron*, **62**, 171–189.

Hall, A.J. & Lomber, S.G. (2008) Auditory cortex projections target the peripheral field representation of primary visual cortex. *Exp. brain Res.*, **190**, 413–430.

Hammond-Kenny, A., Bajo, V.M., King, A.J., & Nodal, F.R. (2017a) Neural Correlates of Multisensory Behavior in the Auditory Cortex. *Assoc Res Otolaryngol Abs*, 147.

Hammond-Kenny, A., Bajo, V.M., King, A.J., & Nodal, F.R. (2017b) Behavioural benefits of multisensory processing in ferrets. *Eur. J. Neurosci.*, **45**, 278–289.

Harrington, L.K. & Peck, C.K. (1998) Spatial disparity affects visual-auditory interactions in human sensorimotor processing. *Exp. brain Res.*, **122**, 247–252.

Hartline, P.H., Vimal, R.L., King, A.J., Kurylo, D.D., & Northmore, D.P. (1995) Effects of eye position on auditory localization and neural representation of space in superior colliculus of cats. *Exp. brain Res.*, **104**, 402–408.

Heffner, H.E. & Heffner, R.S. (1990) Effect of bilateral auditory cortex lesions on sound localization in Japanese macaques. *J. Neurophysiol.*, **64**, 915–931.

Henderson, Z. (1985) Distribution of ganglion cells in the retina of adult pigmented ferret. *Brain Res.*, **358**, 221–228.

Henschke, J.U., Noesselt, T., Scheich, H., & Budinger, E. (2015) Possible anatomical pathways for short-latency multisensory integration processes in primary

sensory cortices. *Brain Struct. Funct.*, **220**, 955–977.

Ho, C., Gray, R., & Spence, C. (2013) Role of audiovisual synchrony in driving head orienting responses. *Exp. Brain Res.*, **227**, 467–476.

Hollensteiner, K.J., Pieper, F., Engler, G., König, P., & Engel, A.K. (2015) Crossmodal Integration Improves Sensory Detection Thresholds in the Ferret. *PLoS One*, **10**, e0124952.

Holmes, N.P. (2007) The law of inverse effectiveness in neurons and behaviour: Multisensory integration versus normal variability. *Neuropsychologia*, **45**, 3340–3345.

Homma, N.Y., Bajo, V.M., Happel, M.F.K., Nodal, F.R., & King, A.J. (2016) Mistuning detection performance of ferrets in a go/no-go task. *J. Acoust. Soc. Am.*, **139**, EL246.

Hwang, J. & Romanski, L.M. (2015) Prefrontal neuronal responses during audiovisual mnemonic processing. *J. Neurosci.*, **35**, 960–971.

Imig, T.J., Irons, W.A., & Samson, F.R. (1990) Single-unit selectivity to azimuthal direction and sound pressure level of noise bursts in cat high-frequency primary auditory cortex. *J. Neurophysiol.*, **63**, 1448–1466.

Itatani, N. & Klump, G.M. (2017) Animal models for auditory streaming. *Philos. Trans. R. Soc. B Biol. Sci.*, **372**, 20160112.

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

Javier, L.K., McGuire, E.A., & Middlebrooks, J.C. (2016) Spatial Stream Segregation by Cats. *J. Assoc. Res. Otolaryngol.*, **17**, 195–207.

Jay, M.F. & Sparks, D.L. (1984) Auditory receptive fields in primate superior colliculus shift with changes in eye position. *Nature*, **309**, 345–347.

Jones, E.G. & Burton, H. (1974) Cytoarchitecture and somatic sensory connectivity of thalamic nuclei other than the ventrobasal complex in the cat. *J. Comp. Neurol.*, **154**, 395–432.

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

Kaltenbach, J.A. (2007) The dorsal cochlear nucleus as a contributor to tinnitus: mechanisms underlying the induction of hyperactivity. *Prog. Brain Res.*, **166**, 89–106.

Kanold, P.O. & Young, E.D. (2001) Proprioceptive information from the pinna provides somatosensory input to cat dorsal cochlear nucleus. *J. Neurosci.*, **21**, 7848–7858.

Katoh, Y.Y. & Benedek, G. (1995) Organization of the colliculo-suprageniculate pathway in the cat: a wheat germ agglutinin-horseradish peroxidase study. *J. Comp. Neurol.*, **352**, 381–397.

Kavanagh, G. & Kelly, J. (1987a) Localization by the Ferret (Mustela putorius). *J. Neurophysiol.*, **57**, 1746–1766.

Kavanagh, G.L. & Kelly, J.B. (1987b) Contribution of auditory cortex to sound localization by the ferret (Mustela putorius). *J. Neurophysiol.*, **57**, 1746–1766.

Kayser, C., Logothetis, N.K., & Panzeri, S. (2010) Visual Enhancement of the Information Representation in Auditory Cortex. *Curr. Biol.*, **20**, 19–24.

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.

Kayser, C., Petkov, C.I., & Logothetis, N.K. (2008) Visual modulation of neurons in auditory cortex. *Cereb. Cortex*, **18**, 1560–1574.

Kayser, C., Petkov, C.I., & Logothetis, N.K. (2009) Multisensory interactions in primate auditory cortex: fMRI and electrophysiology. *Hear. Res.*, **258**, 80–88.

Keating, P., Dahmen, J.C., & King, A.J. (2015) Complementary adaptive processes contribute to the developmental plasticity of spatial hearing. *Nat. Neurosci.*, **18**, 185–187.

Keating, P., Nodal, F.R., & King, A.J. (2014) Behavioural sensitivity to binaural spatial cues in ferrets: evidence for plasticity in the duplex theory of sound localization. *Eur. J. Neurosci.*, **39**, 197–206.

Kelly, J.B. & Judge, P.W. (1994) Binaural organization of primary auditory cortex in the ferret (Mustela putorius). *J. Neurophysiol.*, **71**, 904–913.

Kelly, J.B., Judge, P.W., & Phillips, D.P. (1986) Representation of the cochlea in primary auditory cortex of the ferret (Mustela putorius). *Hear. Res.*, **24**, 111–115.

Keniston, L.P., Allman, B.L., Meredith, M.A., & Clemo, H.R. (2009) Somatosensory and multisensory properties of the medial bank of the ferret rostral suprasylvian

sulcus. *Exp. brain Res.*, **196**, 239–251.

King, A.J. (2002) Neural plasticity: How the eye tells the brain about sound Location. *Curr. Biol.*, **12**, 393–395.

King, A.J. (2009) Visual influences on auditory spatial learning. *Philos. Trans. R. Soc. Lond. B. Biol. Sci.*, **364**, 331–339.

King, A.J. & Hutchings, M.E. (1987) Spatial response properties of acoustically responsive neurons in the superior colliculus of the ferret: a map of auditory space. *J. Neurophysiol.*, **57**, 596–624.

King, A.J., Hutchings, M.E., Moore, D.R., & Blakemore, C. (1988) Developmental plasticity in the visual and auditory representations in the mammalian superior colliculus. *Nature*, **332**, 73–76.

King, A.J. & Palmer, A.R. (1985) Integration of visual and auditory information in bimodal neurones in the guinea-pig superior colliculus. *Exp. brain Res.*, **60**, 492–500.

King, A.J. & Parsons, C.H. (1999) Improved auditory spatial acuity in visually deprived ferrets. *Eur. J. Neurosci.*, **11**, 3945–3956.

King, A.J. & Schnupp, J.W.H. (2000) Sensory convergence in neural function and development. In Gazzaniga, M.S. (ed), *The New Cognitive Neurosciences*, 2nd edn. MIT Press, Cambridge, Massachusetts, pp. 437–450.

King, A.J., Schnupp, J.W.H., & Doubell, T.P. (2001) The shape of ears to come: dynamic coding of auditory space. *Trends Cogn. Sci.*, **5**, 261–270.

Koehler, S.D. & Shore, S.E. (2013) Stimulus-Timing Dependent Multisensory

Plasticity in the Guinea Pig Dorsal Cochlear Nucleus. *PLoS One*, **8**.

Kondo, H.M., van Loon, A.M., Kawahara, J.-I., & Moore, B.C.J. (2017) Auditory and visual scene analysis: an overview. *Philos. Trans. R. Soc. B Biol. Sci.*, **372**, 20160099.

Koppen, C. & Spence, C. (2007) Spatial coincidence modulates the Colavita visual dominance effect. *Neurosci. Lett.*, **417**, 107–111.

Kowalski, N., Versnel, H., & Shamma, S.A. (1995) Comparison of responses in the anterior and primary auditory fields of the ferret cortex. *J. Neurophysiol.*, **73**, 1513–1523.

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**, 279–292.

Lakatos, P., Karmos, G., Mehta, A.D., Ulbert, I., & Schroeder, C.E. (2008a) Entrainment of Neuronal Attentional Selection. *Science (80-.).*, **320**, 23–25.

Lakatos, P., Karmos, G., Mehta, A.D., Ulbert, I., & Schroeder, C.E. (2008b) Entrainment of neuronal oscillations as a mechanism of attentional selection. *Science*, **320**, 110–113.

Laramée, M.E., Kurotani, T., Rockland, K.S., Bronchti, G., & Boire, D. (2011) Indirect pathway between the primary auditory and visual cortices through layer V pyramidal neurons in V2L in mouse and the effects of bilateral enucleation. *Eur. J. Neurosci.*, **34**, 65–78.

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. *Exp. brain Res.*, **158**, 405–414.

Leach, N.D., Nodal, F.R., Cordery, P.M., King, A.J., & Bajo, V.M. (2013) Cortical cholinergic input is required for normal auditory perception and experience-dependent plasticity in adult ferrets. *J Neurosci*, **33**, 6659–6671.

Lee, C.C. & Winer, J.A. (2011) Convergence of thalamic and cortical pathways in cat auditory cortex. *Hear. Res.*, **274**, 85–94.

Leo, F., Bolognini, N., Passamonti, C., Stein, B.E., & Làdavas, E. (2008) Cross-modal localization in hemianopia: new insights on multisensory integration. *Brain*, **131**, 855–865.

Leopold, D.A. & Logothetis, N.K. (1996) Activity changes in early visual cortex reflect monkeys' percepts during binocular rivalry. *Nature*, **379**, 549–553.

Lewald, J., Ehrenstein, W.H., & Guski, R. (2001) Spatio-temporal constraints for auditory--visual integration. *Behav. Brain Res.*, **121**, 69–79.

Lewald, J., Foltys, H., & Töpper, R. (2002) Role of the posterior parietal cortex in spatial hearing. *J. Neurosci.*, **22**, RC207.

Lohse, M., Bajo, V.M., & King, A.J. (2017) Types and distribution of multisensory interactions in auditory thalamus. *Assoc Res Otolaryngol Abs*, 280.

Lomber, S.G. & Malhotra, S. (2008) Double dissociation of "what" and "where" processing in auditory cortex. *Nat. Neurosci.*, **11**, 609–616.

Lomber, S.G., Meredith, M.A., & Kral, A. (2010) Cross-modal plasticity in specific auditory cortices underlies visual compensations in the deaf. *Nat. Neurosci.*, **13**,

1421–1427.

Lomber, S.G., Payne, B.R., & Cornwell, P. (2001) Role of the superior colliculus in analyses of space: superficial and intermediate layer contributions to visual orienting, auditory orienting, and visuospatial discriminations during unilateral and bilateral deactivations. *J. Comp. Neurol.*, **441**, 44–57.

Luo, H., Liu, Z., & Poeppel, D. (2010) Auditory cortex tracks both auditory and visual stimulus dynamics using low-frequency neuronal phase modulation. *PLoS Biol.*, **8**, 25–26.

Ma, L., Micheyl, C., Yin, P., Oxenham, A.J., & Shamma, S.A. (2010) Behavioral measures of auditory streaming in ferrets (Mustela putorius). *J. Comp. Psychol.*, **124**, 317–330.

Maddox, R.K., Pospisil, D.A., Stecker, G.C., & Lee, A.K.C. (2014) Directing eye gaze enhances auditory spatial cue discrimination. *Curr. Biol.*, **24**, 748–752.

Maier, J.X. & Groh, J.M. (2009) Multisensory guidance of orienting behavior. *Hear. Res.*, **258**, 106–112.

Makous, J.C. & Middlebrooks, J.C. (1990) Two-dimensional sound localization by human listeners. *J. Acoust. Soc. Am.*, **87**, 2188–2200.

Malhotra, S. & Lomber, S.G. (2007) Sound localization during homotopic and heterotopic bilateral cooling deactivation of primary and nonprimary auditory cortical areas in the cat. *J. Neurophysiol.*, **97**, 26–43.

Manger, P.R., Engler, G., Moll, C.K.E., & Engel, A.K. (2005) The anterior ectosylvian visual area of the ferret: a homologue for an enigmatic visual

cortical area of the cat? *Eur. J. Neurosci.*, **22**, 706–714.

Manger, P.R., Nakamura, H., Valentiniene, S., & Innocenti, G.M. (2004) Visual areas in the lateral temporal cortex of the ferret (Mustela putorius). *Cereb. Cortex*, **14**, 676–689.

Marmolejo-Ramos, F., Cousineau, D., Benites, L., & Maehara, R. (2014) On the efficacy of procedures to normalize Ex-Gaussian distributions. *Front. Psychol.*, **5**, 1548.

May, B.J. (2000) Role of the dorsal cochlear nucleus in the sound localization behavior of cats. *Hear. Res.*, **148**, 74–87.

McGettigan, C., Faulkner, A., Altarelli, I., Obleser, J., Baverstock, H., & Scott, S.K. (2012) Speech comprehension aided by multiple modalities: Behavioural and neural interactions. *Neuropsychologia*, **50**, 762–776.

McGurk, H. & MacDonald, J. (1976) Hearing lips and seeing voices. *Nature*, **264**, 746–748.

Meijer, G.T., Montijn, J.S., Pennartz, C.M.A., & Lansink, C.S. (2017) Audiovisual Modulation in Mouse Primary Visual Cortex Depends on Cross-Modal Stimulus Configuration and Congruency. *J. Neurosci.*, **37**, 8783–8796.

Meredith, A.M. & Stein, B.E. (1986a) Spatial factors determine the activity of multisensory neurons in cat superior colliculus. *Brain Res.*, **365**, 350–354.

Meredith, M.A. & Allman, B.L. (2015) Single-unit analysis of somatosensory processing in the core auditory cortex of hearing ferrets. *Eur. J. Neurosci.*, **41**, 686–698.

Meredith, M.A., Allman, B.L., Keniston, L.P., & Clemo, H.R. (2012) *Are Bimodal Neurons the Same throughout the Brain?*, The Neural Bases of Multisensory Processes.

Meredith, M.A., Clemo, H.R., Dehner, L.R., & Dehneri, L.R. (2000) Responses to innocuous, but not noxious, somatosensory stimulation by neurons in the ferret superior colliculus. *Somatosens. Mot. Res.*, **17**, 297–308.

Meredith, M.A., Nemitz, J.W., & Stein, B.E. (1987) Determinants of multisensory integration in superior colliculus neurons. I. Temporal factors. *J. Neurosci.*, **7**, 3215–3229.

Meredith, M.A. & Stein, B.E. (1983) Interactions among converging sensory inputs in the superior colliculus. *Science*, **221**, 389–391.

Meredith, M.A. & Stein, B.E. (1986b) Visual, auditory, and somatosensory convergence on cells in superior colliculus results in multisensory integration. *J. Neurophysiol.*, **56**, 640–662.

Meyers, E.M. (2013) The neural decoding toolbox. *Front. Neuroinform.*, **7**, 1–12.

Middlebrooks, J. & Pettigrew, J. (1981) Functional classes of neurons in primary auditory cortex of the cat distinguished by sensitivity to sound location. *J Neurosci*, **1**, 107–120.

Middlebrooks, J.C. & Onsan, Z.A. (2012) Stream segregation with high spatial acuity. *J. Acoust. Soc. Am.*, **132**, 3896–3911.

Miller, J. (1982) Divided attention: evidence for coactivation with redundant signals. *Cogn. Psychol.*, **14**, 247–279.

Miller, L. (1973) Compounding of discriminative stimuli that maintain responding on separate response levers. *J. Exp. Anal. Behav.*, **20**, 57–69.

Miller, L.M. & Recanzone, G.H. (2009) Populations of auditory cortical neurons can accurately encode acoustic space across stimulus intensity. *Proc. Natl. Acad. Sci. U. S. A.*, **106**, 5931–5935.

Molholm, S., Ritter, W., Javitt, D.C., & Foxe, J.J. (2004) Multisensory visual-auditory object recognition in humans: a high-density electrical mapping study. *Cereb. Cortex*, **14**, 452–465.

Morin, L.P. & Studholme, K.M. (2014) Retinofugal projections in the mouse. *J. Comp. Neurol.*, **522**, 3733–3753.

Mrsic-Flogel, T.D., King, A.J., Jenison, R.L., & Schnupp, J.W. (2001) Listening through different ears alters spatial response fields in ferret primary auditory cortex. *J. Neurophysiol.*, **86**, 1043–1046.

Mrsic-Flogel, T.D., King, A.J., & Schnupp, J.W.H. (2005) Encoding of virtual acoustic space stimuli by neurons in ferret primary auditory cortex. *J. Neurophysiol.*, **93**, 3489–3503.

Mrsic-Flogel, T.D., Schnupp, J.W.H., & King, A.J. (2003) Acoustic factors govern developmental sharpening of spatial tuning in the auditory cortex. *Nat Neurosci*, **6**, 981–988.

Murray, M.M. & Wallace, M.T. (2012) *The Neural Bases of Multisensory Processes*, CRC Press/Taylor & Francis, Boca Raton, FL.

Nelken, I., Bizley, J.K., Nodal, F.R., Ahmed, B., Schnupp, J.W.H., & King, A.J.

(2004) Large-scale organization of ferret auditory cortex revealed using continuous acquisition of intrinsic optical signals. *J. Neurophysiol.*, **92**, 2574–2588.

Nelken, I., Chechik, G., Mrsic-Flogel, T.D., King, A.J., & Schnupp, J.W.H. (2005) Encoding stimulus information by spike numbers and mean response time in primary auditory cortex. *J. Comput. Neurosci.*, **19**, 199–221.

Niwa, M., Johnson, J.S., O'Connor, K.N., & Sutter, M.L. (2012a) Active engagement improves primary auditory cortical neurons' ability to discriminate temporal modulation. *J. Neurosci.*, **32**, 9323–9334.

Niwa, M., Johnson, J.S., O'Connor, K.N., & Sutter, M.L. (2012b) Activity Related to Perceptual Judgment and Action in Primary Auditory Cortex. *J. Neurosci.*, **32**, 3193–3210.

Nodal, F.R., Bajo, V.M., & King, A.J. (2012) Plasticity of spatial hearing: behavioural effects of cortical inactivation. *J. Physiol.*, **590**, 3965–3986.

Nodal, F.R., Bajo, V.M., Parsons, C.H., Schnupp, J.W., & King, A.J. (2008) Sound localization behavior in ferrets: comparison of acoustic orientation and approach-to-target responses. *Neuroscience*, **154**, 397–805.

Nodal, F.R., Kacelnik, O., Bajo, V.M., Bizley, J.K., Moore, D.R., & King, A.J. (2010) Lesions of the auditory cortex impair azimuthal sound localization and its recalibration in ferrets. *J. Neurophysiol.*, **103**, 1209–1225.

O'Connor, K.N. & Sutter, M.L. (2000) Global spectral and location effects in auditory perceptual grouping. *J. Cogn. Neurosci.*, **12**, 342–354.

Odegaard, B., Wozny, D.R., & Shams, L. (2015) Biases in Visual, Auditory, and Audiovisual Perception of Space. *PLoS Comput. Biol.*, **11**, e1004649.

Odgaard, E.C., Arieh, Y., & Marks, L.E. (2003) Cross-modal enhancement of perceived brightness: sensory interaction versus response bias. *Percept. Psychophys.*, **65**, 123–132.

Odgaard, E.C., Arieh, Y., & Marks, L.E. (2004) Brighter noise: sensory enhancement of perceived loudness by concurrent visual stimulation. *Cogn. Affect. Behav. Neurosci.*, **4**, 127–132.

Okada, K., Venezia, J.H., Matchin, W., Saberi, K., & Hickok, G. (2013) An fMRI Study of Audiovisual Speech Perception Reveals Multisensory Interactions in Auditory Cortex. *PLoS One*, **8**, 1–8.

Olcese, U., Iurilli, G., & Medini, P. (2013) Cellular and synaptic architecture of multisensory integration in the mouse neocortex. *Neuron*, **79**, 579–593.

Opstal, A.J. van & Munoz, D.P. (2004) Auditory-visual interactions subserving primate gaze orienting. *Calvert, G.; Spence, C.; Stein, B. (eds.), Handb. Multisensory Process.*, 373–394.

Otazu, G.H., Tai, L.-H., Yang, Y., & Zador, A.M. (2009) Engaging in an auditory task suppresses responses in auditory cortex. *Nat. Neurosci.*, **12**, 646–654.

Pallas, S.L. & Sur, M. (1993) Visual projections induced into the auditory pathway of ferrets: II. Corticocortical connections of primary auditory cortex. *J. Comp. Neurol.*, **337**, 317–333.

Park, H., Kayser, C., Thut, G., & Gross, J. (2016) Lip movements entrain the

observers' low-frequency brain oscillations to facilitate speech intelligibility. *Elife*, **5**, 1–17.

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**, 125.

Pesaran, B. (2009) Uncovering the mysterious origins of local field potentials. *Neuron*, **61**, 1–2.

Philipp, R., Distler, C., & Hoffmann, K.P. (2006) A motion-sensitive area in ferret extrastriate visual cortex: An analysis in pigmented and albino animals. *Cereb. Cortex*, **16**, 779–790.

Phillips, D.P., Judge, P.W., & Kelly, J.B. (1988) Primary auditory cortex in the ferret (Mustela putorius): neural response properties and topographic organization. *Brain Res.*, **443**, 281–294.

Pinchoff, R.J., Burkard, R.F., Salvi, R.J., Coad, M.L., & Lockwood, A.H. (1998) Modulation of tinnitus by voluntary jaw movements. *Am. J. Otol.*, **19**, 785–789.

Poole, T.B. (1978) An analysis of social play in polecats (Mustelidae) with comments on the form and evolutionary history of the open mouth play face. *Anim. Behav.*, **26**, 36–49.

Porter, K.K., Metzger, R.R., & Groh, J.M. (2007) Visual- and saccade-related signals in the primate inferior colliculus. *Proc. Natl. Acad. Sci. U. S. A.*, **104**, 17855–17860.

Raab, D.H. (1962) Statistical facilitation of simple reaction times. *Trans. N. Y. Acad.*

Sci., **24**, 574–590.

Rajan, R., Aitkin, L.M., & Irvine, D.R. (1990) Azimuthal sensitivity of neurons in primary auditory cortex of cats. II. Organization along frequency-band strips. *J. Neurophysiol.*, **64**, 888–902.

Randich, A., Klein, R.M., & Lolordo, V.M. (1978) Visual dominance in the pigeon. *J. Exp. Anal. Behav.*, **30**, 129–137.

Reale, R.A. & Imig, T.J. (1980) Tonotopic organization in auditory cortex of the cat. *J. Comp. Neurol.*, **192**, 265–291.

Recanzone, G.H. (1998) Rapidly induced auditory plasticity: the ventriloquism aftereffect. *Proc. Natl. Acad. Sci. U. S. A.*, **95**, 869–875.

Recanzone, G.H., Guard, D.C., & Phan, M.L. (2000) Frequency and intensity response properties of single neurons in the auditory cortex of the behaving macaque monkey. *J. Neurophysiol.*, **83**, 2315–2331.

Recanzone, G.H., Guard, D.C., Phan, M.L., & Su, T.K. (2000) Correlation between the activity of single auditory cortical neurons and sound-localization behavior in the macaque monkey. *J. Neurophysiol.*, **83**, 2723–2739.

Redies, H., Sieben, U., & Creutzfeldt, O.D. (1989) Functional subdivisions in the auditory cortex of the guinea pig. *J. Comp. Neurol.*, **282**, 473–488.

Rockland, K.S. & Ojima, H. (2003) Multisensory convergence in calcarine visual areas in macaque monkey. *Int. J. Psychophysiol.*, **50**, 19–26.

Rodgers, C.C. & DeWeese, M.R. (2014) Neural correlates of task switching in prefrontal cortex and primary auditory cortex in a novel stimulus selection task

for rodents. *Neuron*, **82**, 1157–1170.

Rohe, T. & Noppeney, U. (2016) Distinct Computational Principles Govern Multisensory Integration in Primary Sensory and Association Cortices. *Curr. Biol.*, **26**, 509–514.

Romo, R., Hernández, A., Zainos, A., Lemus, L., & Brody, C.D. (2002) Neuronal correlates of decision-making in secondary somatosensory cortex. *Nat. Neurosci.*, **5**, 1217–1225.

Rouiller, E.M. & Welker, E. (2000) A comparative analysis of the morphology of corticothalamic projections in mammals. *Brain Res. Bull.*, **53**, 727–741.

Rowland, B.A., Quessy, S., Stanford, T.R., & Stein, B.E. (2007) Multisensory Integration Shortens Physiological Response Latencies. *J. Neurosci.*, **27**, 5879–5884.

Roy, S. & Wang, X. (2012) Wireless multi-channel single unit recording in freely moving and vocalizing primates. *J. Neurosci. Methods*, **203**, 28–40.

Ruhland, J.L., Yin, T.C.T., & Tollin, D.J. (2013) Gaze shifts to auditory and visual stimuli in cats. *J. Assoc. Res. Otolaryngol.*, **14**, 731–755.

Sahibzada, N., Dean, P., & Redgrave, P. (1986) Movements resembling orientation or avoidance elicited by electrical stimulation of the superior colliculus in rats. *J. Neurosci.*, **6**, 723–733.

Sarko, D.K., Ghose, D., & Wallace, M.T. (2013) Convergent approaches toward the study of multisensory perception. *Front. Syst. Neurosci.*, **7**, 81.

Schlack, A., Sterbing-D'Angelo, S.J., Hartung, K., Hoffmann, K.-P., & Bremmer, F.

(2005) Multisensory space representations in the macaque ventral intraparietal area. *J. Neurosci.*, **25**, 4616–4625.

Schnupp, J., Nelken, I., & King, A. (2011) Auditory Neuroscience: making sense of Sound. *MIT Press*,.

Schreiner, C.E. & Winer, J.A. (2007) Auditory cortex mapmaking: principles, projections, and plasticity. *Neuron*, **56**, 356–365.

Schroeder, C.E. & Foxe, J. (2005) Multisensory contributions to low-level, "unisensory" processing. *Curr. Opin. Neurobiol.*, **15**, 454–458.

Schroeder, C.E. & Foxe, J.J. (2002) The timing and laminar profile of converging inputs to multisensory areas of the macaque neocortex. *Cogn. Brain Res.*, **14**, 187–198.

Schroeder, C.E., Lakatos, P., Kajikawa, Y., Partan, S., & Puce, A. (2008) Neuronal oscillations and visual amplification of speech. *Trends Cogn. Sci.*, **12**, 106–113.

Sekiyama, K., Kanno, I., Miura, S., & Sugita, Y. (2003) Auditory-visual speech perception examined by fMRI and PET. *Neurosci. Res.*, **47**, 277–287.

Selezneva, E., Scheich, H., & Brosch, M. (2006) Dual time scales for categorical decision making in auditory cortex. *Curr. Biol.*, **16**, 2428–2433.

Sherman, S.M. (2016) Thalamus plays a central role in ongoing cortical functioning. *Nat. Neurosci.*, **16**, 533–541.

Shore, S.E. (2005) Multisensory integration in the dorsal cochlear nucleus: Unit responses to acoustic and trigeminal ganglion stimulation. *Eur. J. Neurosci.*, **21**, 3334–3348.

Shore, S.E. & Zhou, J. (2006) Somatosensory influence on the cochlear nucleus and beyond. *Hear. Res.*, **216–217**, 90–99.

Shuler, M.G. & Bear, M.F. (2006) Reward timing in the primary visual cortex. *Science*, **311**, 1606–1609.

Sinnett, S., Spence, C., & Soto-Faraco, S. (2007) Visual dominance and attention: The Colavita effect revisited. *Percept. {&} Psychophys.*, **69**, 673–686.

Slutsky, D.A. & Recanzone, G.H. (2001) Temporal and spatial dependency of the ventriloquism effect. *Neuroreport*, **12**, 7–10.

Smiley, J.F. & Falchier, A. (2009) Multisensory connections of monkey auditory cerebral cortex. *Hear. Res.*, **258**, 37–46.

Song, Y.-H., Kim, J.-H., Jeong, H.-W., Choi, I., Jeong, D., Kim, K., & Lee, S.-H. (2017) A Neural Circuit for Auditory Dominance over Visual Perception. *Neuron*, **93**, 940–954.e6.

Spence, C. (2007) Audiovisual multisensory integration. *Acoust. Sci. Technol.*, **28**, 61–70.

Stanford, T.R., Quessy, S., & Stein, B.E. (2005) Evaluating the operations underlying multisensory integration in the cat superior colliculus. *J. Neurosci.*, **25**, 6499–6508.

Stecker, G.C., Harrington, I.A., & Middlebrooks, J.C. (2005) Location coding by opponent neural populations in the auditory cortex. *PLoS Biol.*, **3**, 0520–0528.

Stehberg, J., Dang, P.T., & Frostig, R.D. (2014) Unimodal primary sensory cortices are directly connected by long-range horizontal projections in the rat sensory

cortex. *Front. Neuroanat.*, **8**, 93.

Stein, B. (2012) *The New Handbook of Multisensory Processing.* MIT Press, Cambridge.

Stein, B.E., Huneycutt, W.S., & Meredith, M.A. (1988) Neurons and behavior: the same rules of multisensory integration apply. *Brain Res.*, **448**, 355–358.

Stein, B.E., London, N., Wilkinson, L.K., & Price, D.D. (1996) Enhancement of perceived visual intensity by auditory stimuli: a psychophysical analysis. *J. Cogn. Neurosci.*, **8**, 497–506.

Stein, B.E. & Meredith, M.A. (1993) *The Merging of the Senses*. MIT Press, Boston, MA.

Stein, B.E., Meredith, M.A., Huneycutt, W.S., & McDade, L. (1989a) Behavioral Indices of Multisensory Integration: Orientation to Visual Cues is Affected by Auditory Stimuli. *J. Cogn. Neurosci.*, **1**, 12–24.

Stein, B.E., Meredith, M.A., Huneycutt, W.S., & McDade, L. (1989b) Behavioral Indices of Multisensory Integration: Orientation to Visual Cues is Affected by Auditory Stimuli. *J. Cogn. Neurosci.*, **1**, 12–24.

Stein, B.E. & Stanford, T.R. (2008) Multisensory integration: current issues from the perspective of the single neuron. *Nat. Rev. Neurosci.*, **9**, 255–266.

Stein, B.E., Stanford, T.R., Ramachandran, R., Perrault, T.J., & Rowland, B.A. (2009) Challenges in quantifying multisensory integration: alternative criteria, models, and inverse effectiveness. *Exp. brain Res.*, **198**, 113–126.

Stein, B.E. & Wallace, M.T. (1996) Comparisons of cross-modality integration in

midbrain and cortex. *Prog. Brain Res.*, **112**, 289–299.

Steinmetz, P.N., Roy, a, Fitzgerald, P.J., Hsiao, S.S., Johnson, K.O., & Niebur, E. (2000) Attention modulates synchronized neuronal firing in primate somatosensory cortex. *Nature*, **404**, 187–190.

Stevenson, R.A., Ghose, D., Fister, J.K., Sarko, D.K., Altieri, N.A., Nidiffer, A.R., Kurela, L.A.R., Siemann, J.K., James, T.W., & Wallace, M.T. (2014) Identifying and Quantifying Multisensory Integration: A Tutorial Review. *Brain Topogr.*, **27**, 707–730.

Stitt, I., Galindo-Leon, E., Pieper, F., Hollensteiner, K.J., Engler, G., & Engel, A.K. (2015) Auditory and visual interactions between the superior and inferior colliculi in the ferret. *Eur. J. Neurosci.*, **41**, 1311–1320.

Stone, J. (1965) A quantitative analysis of the distribution of ganglion cells in the cat's retina. *J. Comp. Neurol.*, **124**, 337–352.

Stricanne, B., Andersen, R.A., & Mazzoni, P. (1996) Eye-centered, head-centered, and intermediate coding of remembered sound locations in area LIP. *J. Neurophysiol.*, **76**, 2071–2076.

Sumby, W.H. & Pollack, I. (1954) Visual Contribution to Speech Intelligibility in Noise. *J. Acoust. Soc. Am.*, **26**, 212–215.

Sutter, M.L. & Schreiner, C.E. (1991) Physiology and topography of neurons with multipeaked tuning curves in cat primary auditory cortex. *J. Neurophysiol.*, **65**, 1207–1226.

Tan, A.Y.Y., Zhang, L.I., Merzenich, M.M., & Schreiner, C.E. (2004) Tone-evoked

excitatory and inhibitory synaptic conductances of primary auditory cortex neurons. *J. Neurophysiol.*, **92**, 630–643.

Thomas, G.J. (1941) Experimental study of the influence of vision on sound localization. *J. Exp. Psychol.*, **28**, 163–177.

Thomas, H., Tillein, J., Heil, P., & Scheich, H. (1993) Functional organization of auditory cortex in the mongolian gerbil (Meriones unguiculatus). I. Electrophysiological mapping of frequency representation and distinction of fields. *Eur. J. Neurosci.*, **5**, 882–897.

Thorne, J.D., De Vos, M., Viola, F.C., & Debener, S. (2011) Cross-Modal Phase Reset Predicts Auditory Task Performance in Humans. *J. Neurosci.*, **31**, 3853–3861.

Todd, J.W. (1912) Reaction to multiple stimuli. *Arch. Psychol.*, **3**, 1–65.

Tolnai, S., Litovsky, R.Y., & King, A.J. (2014) The precedence effect and its buildup and breakdown in ferrets and humans. *J. Acoust. Soc. Am.*, **135**, 1406–1418.

Tsodyks, M., Kenet, T., Grinvald, A., & Arieli, A. (1999) Linking spontaneous activity of single cortical neurons and the underlying functional architecture. *Science*, **286**, 1943–1946.

Tsunada, J., Liu, A.S.K., Gold, J.I., & Cohen, Y.E. (2015) Causal contribution of primate auditory cortex to auditory perceptual decision-making. *Nat. Neurosci.*, **19**, 135–142.

Van Atteveldt, N.M., Van Aalderen-Smeets, S.I., Jacobi, C., & Ruigrok, N. (2014) Media reporting of neuroscience depends on timing, topic and newspaper type.

PLoS One, **9**.

Vatakis, A., Navarra, J., Soto-Faraco, S., & Spence, C. (2008) Audiovisual temporal adaptation of speech: temporal order versus simultaneity judgments. *Exp. brain Res.*, **185**, 521–529.

Vatakis, A. & Spence, C. (2007) Crossmodal binding: evaluating the "unity assumption" using audiovisual speech stimuli. *Percept. Psychophys.*, **69**, 744–756.

von Melchner, L., Pallas, S.L., & Sur, M. (2000) Visual behaviour mediated by retinal projections directed to the auditory pathway. *Nature*, **404**, 871–876.

Walker, K.M.M., Schnupp, J.W.H., Hart-Schnupp, S.M.B., King, A.J., & Bizley, J.K. (2009) Pitch discrimination by ferrets for simple and complex sounds. *J. Acoust. Soc. Am.*, **126**, 1321–1335.

Wallace, M.N. & Harper, M.S. (1997) Callosal connections of the ferret primary auditory cortex. *Exp. brain Res.*, **116**, 367–374.

Wallace, M.N., Roeda, D., & Harper, M.S. (1997) Deoxyglucose uptake in the ferret auditory cortex. *Exp. brain Res.*, **117**, 488–500.

Wallace, M.T., Meredith, M.A., & Stein, B.E. (1992) Integration of multiple sensory modalities in cat cortex. *Exp. brain Res.*, **91**, 484–488.

Wallace, M.T., Meredith, M. a, & Stein, B.E. (1998) Multisensory integration in the superior colliculus of the alert cat. *J. Neurophysiol.*, **80**, 1006–1010.

Wallace, M.T., Ramachandran, R., & Stein, B.E. (2004) A revised view of sensory cortical parcellation. *Proc. Natl. Acad. Sci. U. S. A.*, **101**, 2167–2172.

Wallace, M.T. & Stein, B.E. (1996) Sensory organization of the superior colliculus in cat and monkey. *Prog. Brain Res.*, **112**, 301–311.

Wallace, M.T., Wilkinson, L.K., & Stein, B.E. (1996) Representation and integration of multiple sensory inputs in primate superior colliculus. *J. Neurophysiol.*, **76**, 1246–1266.

Wang, Y., Celebrini, S., Trotter, Y., & Barone, P. (2008) Visual-auditory interactions in the primary visual cortex of the behaving monkey: electrophysiological evidence. *BMC Neurosci.*, **9**, 79.

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. *Percept. Psychophys.*, **30**, 557–564.

Wasserman, S.M., Aptekar, J.W., Lu, P., Nguyen, J., Wang, A.L., Keles, M.F., Grygoruk, A., Krantz, D.E., Larsen, C., & Frye, M.A. (2015) Olfactory neuromodulation of motion vision circuitry in drosophila. *Curr. Biol.*, **25**, 467–472.

Welch, R.B. & Warren, D.H. (1980) Immediate perceptual response to intersensory discrepancy. *Psychol. Bull.*, **88**, 638–667.

Werner-Reiss, U., Kelly, K.A., Trause, A.S., Underhill, A.M., & Groh, J.M. (2003) Eye position affects activity in primary auditory cortex of primates. *Curr. Biol.*, **13**, 554–562.

Wigderson, E., Nelken, I., & Yarom, Y. (2016) Early multisensory integration of self and source motion in the auditory system. *Proc. Natl. Acad. Sci. U. S. A.*, **113**,

8308–8313.

Wilkinson, L.K., Meredith, M.A., & Stein, B.E. (1996) The role of anterior ectosylvian cortex in cross-modality orientation and approach behavior. *Exp. brain Res.*, **112**, 1–10.

Winer, J. a & Lee, C.C. (2007) The distributed auditory cortex. *Hear. Res.*, **229**, 3–13.

WITKIN, H.A., WAPNER, S., & LEVENTHAL, T. (1952) Sound localization with conflicting visual and auditory cues. *J. Exp. Psychol.*, **43**, 58–67.

Wolpert, D.M., Ghahramani, Z., & Jordan, M.I. (1995) An internal model for sensorimotor integration. *Science*, **269**, 1880–1882.

Yang, T. & Maunsell, J.H.R. (2004) The Effect of Perceptual Learning on Neuronal Responses in Monkey Visual Area V4. *J. Neurosci.*, **24**, 1617–1626.

Yao, J.D., Bremen, P., & Middlebrooks, J.C. (2013) Rat primary auditory cortex is tuned exclusively to the contralateral hemifield. *J Neuro-physiol*, **110**, 2140–2151.

Yin, P., Fritz, J.B., & Shamma, S.A. (2010) Do ferrets perceive relative pitch? *J. Acoust. Soc. Am.*, **127**, 1673–1680.

Yu, J.J. & Young, E.D. (2000) Linear and nonlinear pathways of spectral information transmission in the cochlear nucleus. *Proc. Natl. Acad. Sci. U. S. A.*, **97**, 11780–11786.

Zatorre, R.J., Bouffard, M., Ahad, P., & Belin, P. (2002) Where is “where” in the human auditory cortex? *Nat. Neurosci.*, **5**, 905–909.

Zwiers, M.P., Van Opstal, A.J., & Paige, G.D. (2003) Plasticity in human sound localization induced by compressed spatial vision. *Nat. Neurosci.*, **6**, 175–181.

Zwiers, M.P., Versnel, H., & Van Opstal, A.J. (2004) Involvement of monkey inferior colliculus in spatial hearing. *J. Neurosci.*, **24**, 4145–4156.