The neural foundations of an auditory-visual illusion: spatial factors and application to childhood development.

Hamish Innes-Brown

Submitted for the degree of Doctor of Philosophy

2012

Brain Sciences Institute,
Swinburne University.
0.1 Abstract

Most animals, including humans, possess multiple sense organs through which information about the environment is gleaned. The brain’s ability to integrate the information gathered from multiple sense modalities is fundamental to success in the environment. However, until fairly recently, the processing of sensory information in the brain was assumed to follow separate parallel streams, being progressively elaborated in these separate streams before being finally united in high level association cortices, such as those in the frontal or parietal lobes. As more sophisticated theories and analysis techniques have emerged, this view has given way to a new understanding that information is also transferred between sensory pathways at very early stages – at the level of the primary sensory cortices and possibly below.

The three experiments presented in this thesis used the flash-beep illusion (Shams et al., 2000) as an index of multi-sensory processes in the brain – firstly to show that the illusion was robust to significant spatial separation of the auditory and visual parts of the stimulus, and secondly to show that congruent multi-modal stimuli initiated a sequence of long-lasting effects on subsequent uni-sensory processing that were localised to the parietal and occipital lobes. Lastly, it was shown that a group of children from 8-14 years of age reported significantly more illusions than adults, adding further weight to the growing body of evidence suggesting that many multi-sensory processes are slow to develop.

Overall, the experiments reported in this thesis highlight the involvement of high-level cortical association areas as well as direct cortico-cortical connections between primary auditory and visual areas in the generation of the flash-beep illusion. It is proposed that both are required in order to generate the perception of an illusory flash: the multi-modal flash/beep stimulus initiates a sequence of activity in the primary sensory areas which subsequently leads to feedback from parietal areas, thus priming visual cortex. Direct, spatially insensitive connections from primary auditory to visual areas subserving the visual periphery may allow auditory activity generated by the second beep to trigger a visual percept. The
results increase our understanding that multi-sensory processes are generally the result of complex interactions between sub-cortical nuclei and primary sensory areas, guided by top-down control from high-level association regions which are slow to develop during childhood.
0.2 Acknowledgements.

At the Brain Sciences Institute, Professor David Crewther has always been a source of help when needed, and provided a very vital initial spark of inspiration that led me to move to Melbourne to take up life as a PhD student at the Brain Sciences Institute. The rest of the staff and students at the BSI also deserve thanks, many for participating in long and tiring EEG experiments, and for being subjected to never-ending torrents of 3.5 KHz beeps and small round flashing disks on a computer screen.

Throughout the last phase of completing this thesis I have had the great pleasure of working at the Bionics Institute in Melbourne (known as the Bionic Ear Institute when this research started). The data from Chapter 7 was collected in collaboration with several colleges – Ayla Barutchu in particular deserves special thanks, for helping out at every stage of the research presented in Chapter 7. Associate Professors Antonio Paolini and David Grayden also provided vital support. Throughout the writing-up stage I have also been supported by Dr Jeremy Marozeau and Professor Peter Blamey at the Bionics Institute, who have always provided advice and the space to juggle the complex time and life requirements of writing up a PhD at the same time as working. It is particularly pleasing to have had the opportunity to apply some of the things learnt during the writing up of this PhD to my professional life.

Finally, my biggest thanks go to my wife Kirrily, who has endured the ups and downs of this journey just as much as me, and our very new son Quin, who has been a boundless source of joy through the last phases of writing up this thesis.
0.3 Declaration

I declare that this thesis does not incorporate, without written acknowledgement, any material that has previously been submitted for the award of any other degree or diploma in any university, college, or other educational institution; and to the best of my knowledge, this thesis does not contain any material previously published or written by another person except where due reference is made in the text, including the disclosure of contributions for any work based on joint research or publication.

This thesis contains two published manuscripts and one unpublished experimental chapter, in addition to the introductory and discussion chapters. Chapters 5 and 7 have been published in peer-reviewed journals. Under the agreements for publication, the copyright for the material in Chapter 5 remains with the author. The material in Chapter 7 is re-used here with the permission of the publisher (see Appendix 2). The development, data collection, analysis, interpretation, and write-up of all experimental work in this thesis was the primary responsibility of the author. However, the co-authors for Chapters 5 & 7 contributed to many phases of the research as part of an active scientific collaboration.

I declare that the ethical principles and procedures specified in the Swinburne University of Technology Human Research Ethics document on human research and experimentation have been adhered to in the presentation of this thesis.

Hamish Innes-Brown

Date:
0.4 Table of contents

0.1 Abstract ii
0.2 Acknowledgements. iv
0.3 Declaration v
0.4 Table of contents vi
0.5 List of figures. ix
0.6 List of Tables. xiii
0.7 Publications xiv
0.7.1 Peer-reviewed journal articles reported in this thesis xiv
0.7.2 Other peer-reviewed journal articles xiv
0.7.3 Conference abstracts xv
0.7.4 Lay articles. xviii

1 OPENING 1-1
1.1 Introduction 1-1
1.2 Structure of the thesis 1-2

2 MULTISENSORY INTEGRATION IN THE HUMAN BRAIN 2-5
2.1 Multisensory integration 2-6
2.2 The behavioural consequences and neural mechanisms of multisensory integration 2-7
2.2.1 Auditory effects on visual perception 2-9
2.3 Neural pathways for auditory, visual and auditory-visual processing 2-10
2.3.1 The auditory system 2-10
2.3.2 The visual system 2-15
2.3.3 Auditory-visual integration 2-17
2.4 Electrophysiological correlates of multisensory integration 2-26
2.5 Development 2-28
2.5.1 Matching and integration 2-29
2.6 Conclusion 2-30

3 THE “SOUND-INDUCED FLASH ILLUSION” 3-32
3.1 Perceptual Illusions 3-33
3.3 Electrophysiological correlates 3-42
3.4 Sound-induced fusion illusions 3-47
3.5 fMRI shows V1 activity is associated with the illusory flash 3-48
3.6 The lasting effect of multi-sensory processes 3-50
3.7 Further psychophysics (2007) 3-54
3.8 Advanced Electrophysiology (2007-2009) 3-57
3.9 Applications – Development (2007) 3-58
3.10 General Discussion 3-60

4 EEG ANALYSIS TECHNIQUES 4-61
4.1 Discovery of scalp potentials 4-62
4.2 Generation of the EEG 4-68
4.3 Auditory ERPs 4-70
4.4 Visual ERPs 4-72
4.5 Multisensory (auditory-visual) ERPs 4-73
0.5 List of figures

Figure 2-1. The main ascending (solid lines) and descending (dashed lines) auditory pathways from the left cochlea to the primary auditory cortex. Figure and original caption adapted from (Brugge and Howard, 2003). AC – auditory cortex; MGB – medial geniculate body; IC – inferior colliculus; NLL – nucleus of lateral lemniscus; MNTB – medial nucleus of the trapezoid body; LSO – lateral superior olivary nucleus; MSO – medial superior olivary nucleus; PO – pre-olivary nucleus; DCN – dorsal cochlear nucleus; VCN – ventral cochlear nucleus; AN – auditory nerve (8th cranial nerve).2-12

Figure 2-2. Diagram showing the location of the primary and secondary auditory (A1, A2) and visual (V1, V2) cortex in the human brain. V1 and V2 are located bilaterally on the occipital poles. Although only V1 and V2 are highlighted, many areas within the dotted line are known to be associated with visual processing. A1 and A2 are only partly shown on this view, as they are located partly within the superior temporal sulcus, on the superior surface. Brain outline from (Rosa, 2003), functional regions added after consultation with the Brodmann interactive atlas (Bernal and Perdomo).2-14

Figure 2-3. The projections from the retina through the optic chiasm, LGN, and optic radiation to the primary visual cortex. Figure adapted from (Falchier et al., 2002).2-22

Figure 2-4. An example of non-linear response enhancement for a multisensory neuron in the monkey SC. The top panel shows the visual and auditory spatial receptive fields for the neuron. The middle panel shows raster plots and histograms of the spiking activity of the neuron. The bar graph to the right shows the mean spikes per trial, and indicates that the number of responses in the visual + auditory condition was significantly higher than in either the auditory or the visual condition, and also higher than the sum of the auditory and visual conditions (indicated by the dashed line). Figure with original caption from Wallace, Wilkinson & Stein (1996).2-19

Figure 2-5. The proportion of neurons in the visual cortex (at eccentricities from 0-20 degrees) with direct connections to auditory cortex (light bars) and the polysensory area of the superior temporal lobe (STP – black bars). Note that the proportion of neurons with auditory connections increases with increasing eccentricity, and that the overall proportions are less than that received from STP. Figure and original caption from Falchier et al (2002).2-23

Figure 2-6. Diagram showing the location of the primary and secondary auditory (A1, A2) and visual (V1, V2) cortex in the human brain. V1 and V2 are located bilaterally on the occipital poles. A1 and A2 are only partly shown on this view, as they are located partly within the superior temporal sulcus, on the superior surface. Grey areas from light to dark shades show progressively higher levels of the visual processing stream. A) ‘traditional’ view of the integration of auditory and visual information, where all sensory information is progressively elaborated throughout uni-sensory pathways before being integrated in higher ‘association cortex.’ B) Strictly cortico-cortical connections only. C) Emerging view where, in addition to the flow of information to the higher areas of cortex (grey lines), there are also low-level connections directly between primary sensory cortices. The thick line indicates that evidence has been found for more numerous connections between the primary sensory areas than between secondary areas. Reciprocal connections between A1-A2 and V1-V2 are presumed to exist in both models but are excluded for clarity. *A1 – V1: (Rockland and Ojima, 2003, Falchier et al., 2002, Clavagnier et al., 2004). **A1-V2 (Rockland and Ojima, 2003) ***V2 -> A1: (Budinger et al., 2006).2-25

Figure 3-1. Average number of flashes perceived in Shams et al (2002). The red circle indicates the point where 1F2B, 2F1B and 2FOB data points are overlayed. The points are indistinguishable, indicating that responses to two real flashes alone, or two flashes with two beeps, are similar to when only a single flash is presented, along with two beeps. Figure and original caption adapted from Shams et al (2002).3-41

Figure 3-2. Event-related potentials from Shams et al (2001). Significant multisensory interactions (1F2B – (1FOB + 0F2B)) are present at three occipital electrodes (Oz, O1, O2) from approximately 170-300 ms, only when the fission illusion stimulus is presented in the visual periphery. The top row shows responses when the stimuli were presented in the fovea, the middle row when the stimuli were presented in the visual periphery. The lowest row shows responses to a single flash subtracted from a
double flash (ie leaving only responses to the second flash). Modified figure with original caption from Shams et al. (2001). ........................................................................................................................................3-44

Figure 3-3. Time-frequency plots (average energy, inter-trial coherence and an interaction plot) for illusory vs non-illusory reports for 1F2B trials. Figure and original caption from Bhattacharya et al (2002). ........................................................................................................................................3-45

Figure 3-4. Statistical tests of gamma-band activity (compared to the pre-stimulus baseline) for illusory vs non-illusory reports for 1F2B trials. Figure and original caption from Bhattacharya et al (2002). ........................................................................................................................................3-46

Figure 3-5. Behavioural results from Meylan & Murray (2007). Accuracy was very high in all trial types except the ‘illusory’ 1F2B trials. Figure from Meylan & Murray (2007). ........................................3-51

Figure 4-1. An example of Berger Rhythms following the rate of a flicker stimulus. Figure and original caption from Adrian & Matthews (1934), Brain, 57(4), pp 379. .................................................................4-64

Figure 4-2. One of the first auditory evoked potentials recorded from a human. Figure and original caption from (Davis, 1939). ........................................................................................................................................4-65

Figure 4-3. averages from 55 trials made with the electro-mechanical averager. On the left are the individual trials from an electrical stimulus to the left wrist, and on the right is a 5 μV calibration pulse, again with 55 trials. The top panels show superimposed trials, and the bottom panels the averaged response. Figure and original caption from Merton & Morton (1984). .......................4-66

Figure 4-4. The generation of scalp EEG. From Luck and Girelli (1998). Note that the location of the dipole (immediately below the vertex) is very different to the scalp distribution (skewed to the right), because the dipole is located within a sulcal fold, and thus not oriented tangentially to the scalp. Because of this, scalp maps are not necessarily a good indicator of neural sources. 4-69

Figure 4-5. A cartoon representation of the various component waveforms of the auditory event-related potential. Waves I-VI are generated in the brainstem. Waves P1/P50, N1 and N2 are generated in the cortex but are usually considered ‘obligatory,’ or mainly stimulus-driven, and N2 onwards are ‘endogenous,’ or more related to cognitive events. Figure from (Rissling and Light, 2010). ........................................................................................................................................4-71

Figure 4-6. Time-courses of all independent components calculated for subject N019 in Chapter 6. Components 1, 4, 10, 14, 29, and 40 were marked for rejection in this case. Component 1 appears to be related to blinks, and the remaining marked components are related to occasional electrode problems or high frequency (50 Hz) noise.................................4-76

Figure 4-7. The scalp map, ERP image, and power spectrum of component 1 for subject N019. This component was marked for rejection as it appears to be related to eye-blink activity. .........................4-77

Figure 4-8. An example of the relation between the average ERP and GFP. The top panel shows a butterfly plot of all 60 channels from a combined audio-visual stimulus consisting of a simple visual flash and auditory beep (data taken from the experiments reported in Chapter 6). The bottom panel shows the GFP calculated from the same data. It would be difficult to determine which peaks or parts of the waveform are related to the auditory and visual stimuli – rather, the whole response is a mixture of the two. As can be seen, the peaks in the GFP correspond with periods of maximum activity in the ERP. However, as the measure is reference-free, no decision about electrode references is needed to find the points of maximum activity. ........................................................................4-78

Figure 5-1. Stimulus timing for flashes and beeps. A 2-flash 2-beep (2F2B) trial is shown here........5-90

Figure 5-2. Mean responses for the number of flashes reported (bars show SEM) for all trial types in experiment one. Single-flash trials are shown in panel A, and multi-flash trials in panel B. The reference line in panel B shows veridical perception. ........................................................................5-92
Figure 5-3. Means response rates (%) are shown for both experiment one (grey bars) and experiment 2 (coloured bars). Unisensory, fission- and fusion-capable stimuli are marked with italic text.

Figure 5-4. Mean responses for the number of flashes reported (bars show SEM) for all trial types in experiment two. Single-flash trials are shown in panel A. Multi-flash trials with zero beeps and one beep are shown in panels B and C respectively.

Figure 6-1. Flash beep stimulus timing. Shown is an example of a 2-flash, 2-beep (2F2B) trial.

Figure 6-2. Averaged responses (± SEM) for each stimulus type in one-flash trials (left) and two-flash trials (right).

Figure 6-3. Mean accuracy measures (± SEM) for each stimulus type. In fission illusion trials (1F2B), participants often reported more flashes than were present; in fusion illusion trials (2F1B), participants often reported less flashes than were presented.

Figure 6-4. Example of statistical thresholding technique used to determine bad electrodes. In this example, electrodes 6 and 28 were marked for rejection, as they had mean power outside ± 2 standard deviations from the mean of all electrodes.

Figure 6-5. Diagram showing the derivation of UNI and MULTI difference waves from the uni-modal (1F0B and 2F0B) and MULTI (2F1B and 1F1B) stimulus types.

Figure 6-6. Grand average ERPs for UNI (top panel) and MULTI (bottom panel) stimuli. Significant differences between the waveforms are marked with grey boxes (permutation test, p < .05). The x-axis represents time relative to the first flash.

Figure 6-7. Grand averages for UNI (red) and MULTI (green) difference waves. Grey boxes show time points at which a permutation test indicated a significant difference between the waveforms. The x-axis represents time relative to the first flash.

Figure 6-8. Results from statistical testing of UNI vs MULTI difference waves. The x-axis on all plots represents time relative to the first flash. The grey line at 67 ms indicates the time of the second flash. TOP: P-values from point-wise permutation testing between the UNI and MULTI difference waves at every electrode. Electrodes are arranged in bands separated by grey horizontal lines from the frontal (top of plot) to occipital (bottom of plot) regions. Within each band, electrodes are arranged from left-most (top of band) to right-most (bottom of band). The midline electrode in each band is shown on the y-axis. For example, in the lowest (occipital) band, the three electrodes shown are O1, Oz and O2. MIDDLE: The mean (across participants) global field power (MGFP) calculated across all electrodes for UNI and MULTI difference waves. The MULTI waves shows increased activity in early and late time windows. BOTTOM: p-values from a point-wise permutation test between the two GFP waveforms show that the MULTI difference waveform is significantly larger for a very short period at around 50 ms, and extended periods from 100-200 ms, and from 300-350 ms.

Figure 6-9. Topographic maps of the UNI and MULTI difference waves in the early (top row) and late (bottom row) time intervals. Electrodes at which the mean voltage in each range was statistically significantly different in each condition are indicated with a red dot (p < .05, df = 10).

Figure 6-10. sLORETA source localisation results (µA/mm$^3$) displayed on the MNI T2 template brain for the MULTI (top row) and UNI (bottom row) waves, in the late interval (300-320 ms). The difference between the two (MULTI-UNI) is shown in the bottom row. Four views are shown; from left to right these are top, back, left, and right views.

Figure 6-11. sLORETA statistical image showing significant pseudo-F values from a voxel-wise random permutation test between MULTI and UNI sLORETA source localisations in the early timeframe (130-160 ms, or 63-93 ms after the second flash).

Figure 6-12. sLORETA statistical image showing significant pseudo-F values from a voxel-wise random permutation test between MULTI and UNI sLORETA source localisations in the late timeframe (300-320 ms, or 233-253 ms after the second flash).

Figure 6-13. Timing diagram showing various responses to a simultaneous flash/beep stimulus at time zero. Some points to be identified further below are in response to a difference between stimuli, in which case there is also a flash at 67 ms (for the current study), or a beep at 67 ms (Shams et al
2001, 2005, Mishra et al 2007), in which case there is also an ‘illusory flash’ at approximately 120-130 ms. The green and red circles of points 1 and 2 show the times at which activity to auditory (green circles) and visual (red circles) stimuli first appears in the primary auditory (A1) and visual (V1) areas. Point 3 shows when the activity from the second flash, isolated by the MULTI waveform (2F1B – 1F1B), could be expected to occur. Note that in this case, auditory activity from the initial flash/beep is subtracted and would not appear, leaving only point 3. The green and red inverted triangles (points 4, 5 and 6) show when uni-modal activity from a combined flash/beep at 0 ms reaches association areas in the parietal lobes (PL - 4,5), and when super-additive responses to audio-visual stimuli are found in the same areas shortly afterwards (mixed green and red inverted triangles, 6). The blue boxes show the timing of increased activity found in the current study: points 7, 8: ‘early’ and ‘late’ intervals, respectively, at which MULTI response was significantly larger than the UNI response in the current study. Points 9 and 10 (pink triangles) show increased activity in V1 in participants who were pre-disposed to illusion (point 9), and increased negativity in A1, in illusion compared to non-illusion trials (point 10), in Mishra et al (2007). Point 11 shows the decreased GFP localised to V1 in MULTI compared to UNI responses in Meylan and Murray, (2007). Points 12 & 13 show interactions (1F2B > 1F0B+0F2B) found in parietal (12) and occipital (13) MEG sensors in Shams et al (2005a). Points 14 & 15 show significant interactions (1F2B > 1F0B+0F2B) at occipital EEG electrodes in Shams et al (2001). Times for points 1-3 were extrapolated from primate electrophysiology data to human-equivalent times using ‘3/5ths rule’ (Musacchia and Schroeder, 2009). Details: 1: first onset to broadband noise stimulus in macaque A1 (Lakatos et al., 2005). 2, 3: response to first and second light flash in macaque V1 (Chen et al., 2007). 4, 5, 6: first onset time of activity to auditory (4), visual (5), and integrative response to audio-visual stimulus (6) in human parietal lobe (Moholm et al., 2006). Note there are some small timing differences between the paradigms in Meylan & Murray (2007), who used an ISI of 52 ms; Mishra et al (2007), 70 ms; Shams et al (2005, 2001), 67 ms; and the current study, also 67 ms. Full paradigm details in Table 3-1 in Chapter 3. .................................................................................................................................................................................. 6-145

Figure 6-14. A series of topographic maps spanning the timeframes in which there was a significant difference between the GFP for the UNI (top row) and MULTI conditions (indicated by the blue bars on the x-axis), as the non-significant period in-between. Each map was calculated by taking the average over a 20 ms interval. ........................................................................................................................................... 6-156

Figure 7-1. Flash beep stimulus timing. Shown is an example of a 2-flash, 2-beep (2F2B) trial..... 7-165

Figure 7-2. Mean accuracy measures and 95% confidence intervals (CI) for each stimulus type for adults and children. .................................................................................................................................................. 7-168

Figure 7-3. Reaction times (ms) and 95% CIs for non-illusion (A) and illusion (B) trials. Reaction times for correct responses only are shown in A. In B, reaction times for correct (C) and incorrect (I) responses are shown separately. ..................................................................................................................................... 7-170

Figure 7-4. Scatter plots showing raw accuracy scores for adults and children to all stimuli. Illusion trials are shown on the far right (top and bottom rows). .................................................................................................................. 7-172

Figure 8-1. A) Diagram showing the location of the primary and secondary auditory (A1, A2) and visual (V1, V2) cortex in the human brain. V1 and V2 are located bilaterally on the occipital poles. A1 and A2 are only partly shown on this view, as they are located partly within the superior temporal sulcus, on the superior surface. B) ‘traditional hierarchical’ view of the integration of auditory and visual information, where all sensory information is progressively elaborated throughout uni-sensory pathways before being integrated in higher association cortex. C) Strictly cortico-cortical connections only D) Emerging view where, in addition to the flow of information to the higher areas of cortex (grey lines), there are also low-level connections directly between primary sensory cortices. The thick line indicates that evidence has been found for more numerous connections between the primary sensory areas than between secondary areas. Reciprocal connections between A1-A2 and V1-V2 are presumed to exist in both models but are excluded for clarity. * A1 -> V1: (Rockland and Ojima, 2003, Falchier et al., 2002, Clavagnier et al., 2004). ** A1-V2 (Rockland and Ojima, 2003) *** V2-A1: (Budinger et al., 2006). ........................................................................................................................................ 8-184

Figure 8-2. Three possible models of the generation of the illusory flash in the flash-beep illusion paradigm. Green and red circles represent the first onset of activation in the auditory (A1) and visual (V1) cortices by auditory (green circle, points 1 & 3) and visual (red circle, point 2) stimuli. Green
and red inverted triangles show projected activity in the parietal lobe (PL) associated with the unimodal stimuli. Overlayed green and red triangles show integrative response in the PL. Panel A: ‘traditional hierarchical’ model, where information from A1 and V1 is only combined in high-level parietal association areas. Panel B: ‘cortico-cortical only’ model, where information is combined directly between A1 and V1, with no involvement of high-level cortical areas. Panel C: ‘richly-connected’ model, where both cortico-cortical connections as well as feedback from high-level association areas are involved. References: 1, 3: first onset of activity to broadband noise stimulus in macaque A1 (Lakatos et al., 2005). 2: response to light flash in macaque V1 (Chen et al., 2007). 5, 6, 7: first onset time of activity to auditory (5), visual (6), and integrative response to audio-visual stimulus (7) in human parietal lobe (Molholm et al., 2006).

0.6 List of Tables

Table 3-1. Summary of flash-beep illusion studies. Basic stimulus parameters and results..............3-37
Table 5-1. The odds for fusion and fission illusions for each of the four illusion-capable stimuli in Experiment One (central presentation). The final odds ratio, and the significance of the association between stimulus type (uni- vs. multi-sensory) and reports of illusions (fissions/fusions vs non-illusory) are also shown..............................................................................................................................5-95
Table 5-2. The odds of fusion and fission illusions for multisensory stimuli and their uni-sensory equivalents in combined data from experiments 1 and 2. The LL and RR spatial configurations were collapsed with the data from experiment one. The rows labeled “All fusion” and “All fusion” show data summed across all fission- and fusion-capable stimuli, respectively. The final odds ratio, and the significance of the association between stimulus type (uni- vs. multi-sensory) and reports of illusions (fissions/fusions vs non-illusory) are also shown..............................................................................................................................5-97
Table 6-1. Stimulus-response contingencies. For non-illusion stimuli (1F0B, 1F1B, 2F0B, 2F2B), incorrect responses were not analysed. For illusion stimuli (1F2B, 2F1B), separate codes were assigned depending on the response.............................................................................................................................. 6-117
Table 6-2. Number of trials remaining for each participant in each stimulus/response combination. Most participants had insufficient trials in either ‘illusion’ or ‘no-illusion’ responses to fusion (1F2B) stimuli to allow a direct comparison of illusion vs non-illusory responses to the same stimulus. .......................................................................................................................................................... 6-121
Table 6-3. Locations of the maximum difference (maximum pseudo-F statistic) in the early and late time intervals. .......................................................................................................................................................... 6-140
Table 6-4. Details of flash-beep illusion studies finding interactions in the parietal lobes, A1 and V1.6-152
Table 6-5. Stimulus configuration differences between the current study and Meylan & Murray (2007). .......................................................................................................................................................... 6-154
Table 7-1. Means, standard errors of the mean, and t-test outcomes for Purdue Pegboard Scores.......7-167
Table 7-2 Correlations Between Accuracy Scores for Fission and Fusion Stimuli, Purdue Pegboard Scores, and Reaction Times (RT) for Correctly and Incorrectly Responded Stimuli..........................7-173
0.7 Publications

0.7.1 Peer-reviewed journal articles reported in this thesis


0.7.2 Other peer-reviewed journal articles


**0.7.3 Conference abstracts**


Michie, P. T., **Innes-Brown, H.**, Todd, J., Wynne, B., Frima, R., Dedman, K., et al. (2000). Reduced duration MMN in first degree relatives of patients with a


0.7.4 Lay articles.

1 Opening

1.1 Introduction

The ability to integrate information from more than one sense modality offers great behavioural advantages. In everyday life, our behaviour is constantly guided by a continuous stream of sensory impressions that are built from not only a linear mixture, but an active integration of sensory information from multiple sense modalities. Musicians, dancers, athletes, and even an individual crossing a busy street must combine the redundant but sparse information they receive from individual sense modalities in order to perform either at the limits of behavioural possibility (in the case of musicians etc), or in a safe and reliable manner in everyday life.

Within the last 10-15 years, the study of the mechanisms by which the brain integrates information from the senses has undergone a paradigm shift. Traditional hierarchical models of sensory processing hold that information from each sense modality is processed in separate modality-specific pathways, until finally being united in high-level association areas. This long-held view has only recently given way to a new understanding that information is also transferred between sensory pathways at very early stages, at the level of the primary sensory areas and perhaps sub-cortically.

Although research in this field covers all the sense modalities, the scope of this thesis is restricted to consideration of the effects of auditory stimuli on visual perception. Elements of this new view of multisensory processes were explored using the ‘flash beep illusion’ – an auditory-visual illusion where auditory stimuli have been found to influence visual perception. Three experiments were performed. Firstly, fundamental spatial properties of the illusion were explored using psychophysical techniques. Secondly, the possible neural sources of the illusion were determined using source imaging of the EEG in adults. Thirdly, a translational study used performance on the illusion task to index the development
of multisensory processes in school-aged children. The results from these three studies are brought together in the final chapter under a common framework describing how various networks of brain areas and connectivity may explain the generation of this illusion. A hybrid model combining feedback influences from traditional hierarchical models with direct cortico-cortical connections is proposed to best fit the results.

### 1.2 Structure of the thesis

Chapters 2 and 3 review existing research on multisensory integration and the flash-beep paradigm. Chapter 2 provides a historical review of the literature on multisensory integration. The review covers early experiments in humans describing the facilitatory effect of integrating sensory information, animal studies describing the existence and properties of multisensory neurons in subcortical structures, and later experiments in humans and animals suggesting that multisensory processes also occur early in primary sensory cortex, as well as directly between primary sensory areas. Finally, the development of these processes in human infants and children is reviewed. An ever-increasing number of results from flash-beep illusion paradigms have been published during the writing of this thesis. Chapter 3 will focus specifically on comprehensively reviewing the sometimes-variable results of these behavioural and neuroimaging studies. The impact of experimental methods and analytic procedure is discussed, and the results from the small number of studies employing the illusion in children are discussed.

After describing how the analysis of event-related potential components is not sufficiently sensitive to pick up the dynamic changes in brain activation that may be occurring in the brain during multisensory integration, Chapter 4 briefly review the problems with analysis of event-related potential components, and introduces the alternative methods that will be used in the electrophysiological experiments presented in Chapter 6.
Chapters 5, 6 and 7 present results from three experiments. Chapter 5 describes the initial behavioural replication (in 9 adults) of the flash-beep phenomenon in our laboratory. This was an essential step, as at the time the experiment was performed, there was only a single report of the phenomenon in the literature. After finding the phenomenon was replicable, an extension of the experiment showed how spatially separating the auditory and visual parts of the stimuli had no effect on the strength of the illusion. The results are discussed in terms of emerging theories proposing direct cortico-cortical links supporting the flow of information from auditory to visual cortex. Chapter 5 is a reproduction of a published manuscript (Innes-Brown and Crewther, 2009).

After replicating and exploring the spatial properties of the flash-beep phenomenon in our lab, Chapter 6 presents an electrophysiological examination (in 14 adults) of the brain sources that might support the illusion. Using event-related potentials (ERPs) and source localisation techniques, it was shown that multisensory interactions have a long-lasting effect – an initial multi-modal flash/beep stimulus was found to affect the subsequent processing of a flash stimulus. These effects were localised to two separate timeframes and locations in the brain: an early timeframe, from 130-160 ms after the initial flash/beep pair, where effects were localised to parietal regions, and a later timeframe, from 300-320 ms after the initial flash/beep pair, localised to primary and secondary visual cortex in the occipital lobes. Again, the results were supportive of the view where rapid, direct communication between primary sensory areas is possible, but also highlight the involvement of feed-back connections from parietal association areas. The results from Chapter 6 were also displayed in a new timeline-based framework that described possible brain networks underlying the enduring multisensory effects of the initial flash/beep stimulus on subsequent uni-modal processing.

The methods explored in Chapters 5 and 6 were then applied in Chapter 7 to explore the developmental time-course of multisensory integration. A population of 30 children aged 8-11 was compared with 22 adults. The results showed that the illusion occurred more frequently in the children compared to the adults. These findings are consistent with ideas of neural pruning during
development, and show that multisensory processes do not fully develop until relatively late in childhood. The involvement of late-developing cortical areas on the control of multisensory processes is discussed. Chapter 7 is also a reproduction of a published manuscript (Innes-Brown et al., 2011). EEG data was also recorded for both the adults and children in this group, but the data is not reported here. A preliminary analysis was prepared for a conference poster presentation, however, and the poster is attached in Appendix 2.

Chapter 8 positions the results from Chapter 5-7 within an extended timeline based framework similar to that proposed in Chapter 6. The results from the three chapters are discussed in terms of three possible brain network models: A) traditional hierarchical, B) cortico-cortical connections only, and C) a hybrid model with elements from A and B. It is proposed that model C best explains the results from Chapters 5-7, and can most parsimoniously predict the results from other research employing the flash-beep illusion task.

Overall, the results from the three experimental chapters are consistent with the emerging idea that multisensory processes are more widespread across a range of timescales and brain areas, than the traditional, parallel-hierarchical model allows.

Appendix 1 contains participant information and consent forms.

Appendix 2 contains cover pages of the two published papers that have arisen from this research so far, a copy of the agreement from the publisher to reproduce Innes-Brown et al (2011) as Chapter 7 (permission was not required for Innes-Brown et al, 2009, as the copyright remains with the authors) as well as two selected posters from conference presentations that show an alternative analysis of the EEG data in adults from Chapter 5, and an analysis of the EEG data from the children in Chapter 7.
2 Multisensory integration in the human brain
2.1 Multisensory integration

Humans, like most other animals, possess multiple sensory epithelia through which we perceive the environment. Light energy is transduced into patterns of electrical nerve impulses by the retina, sound pressure waves are transduced by the cochlea, and mechanical forces on the skin and joints are transduced by the somatosensory system. Events or objects in the environment often produce stimuli in more than one domain (for instance both light and sound stimuli are involved when a nearby person claps their hands), and thus more than one sense modality is often stimulated by environmental events. Signals from all the senses are not only brought together in the brain in order to form a unified percept of the world, but are also actively integrated. Perceptual information from one sense is able to improve or alter perception in another, so that the integration of information from multiple senses forms a ‘new product,’ or a unified multisensory experience, probably different from the sum of its parts, which is vital for cognition and the guidance of actions (Stein et al., 2010). Ultimately, an understanding of how the brain achieves these tasks may be important for the development of cybernetic sensory systems (Hall et al., 2009), or in maximising the benefits that sensory prostheses can provide (Innes-Brown et al., 2010, Innes-Brown et al., 2009a, Innes-Brown et al., 2009b).

It is important to clearly define the scope of the term “multisensory integration” (MSI), as used throughout this thesis, as in different parts of the literature the term has come to have quite different meanings depending on the field of study. In the last ten to fifteen years there has been an explosion of research drawing together investigators from many different disciplines, all interested in how the integration of information from different senses affects sensation, perception, and behaviour. Each field carries with it a particular set of assumptions, terminology, and research methods, and this has led to some confusion in the field. As we will see in Section 2.3.3, MSI has a unique ‘super-additive’ (or otherwise non-linear) neural signature in many parts of the brain – a multisensory neuron is sometimes said to be integrating when its response given a
multimodal stimulus is larger than the sum of the responses from each uni-modal input alone. The output of the multisensory neuron is thus a ‘new product’ derived from a complex non-linear interaction among the neurons inputs (Stein et al., 2010). This new neural product is presumed to then lead on to behavioural outcomes, for instance reaction times or detection rates, that are correspondingly ‘super-additive,’ or superior to those realised with either of the unimodal stimuli alone. In this thesis, and following the scheme set out in Stein et al (2010), the suffix ‘-modal’ will refer to features of stimuli, and the suffix ‘-sensory’ will refer to neural processes. For instance, ‘unimodal’ refers to a stimulus with features limited to a single sense modality, and ‘unisensory’ refers to neural or behavioural processes associated with a single sense modality. The term ‘multisensory process’ will be used as a general descriptor of multisensory phenomena or outcomes, and the more specific term ‘multisensory integration’ (MSI) will be used to indicate the previously described new product definition. Other terms such as ‘cross-modal matching,’ which will appear in Section 2.5 on development, relate to multisensory processes comparing each part of a multimodal stimulus, and should not be confused with the more specific MSI.

This rest of this chapter will expand on the concept of MSI, and provide reviews of the brain systems thought to be involved in the process and how they develop. As the later experimental chapters are concerned with the integration of auditory and visual signals, these reviews will mainly focus on these systems. Some of the perceptual effects of MSI are discussed, particularly the effect of auditory stimuli on visual sensation. The basic auditory and visual sensory pathways are briefly reviewed, the roles of midbrain (SC) and cortical areas in MSI are discussed, and the development of multisensory processes in children will be reviewed.

2.2 The behavioural consequences and neural mechanisms of multisensory integration

The human central nervous system has evolved in an environment where external events are usually multimodal in nature. Information from multiple senses
must be bound together when originating from a single object, but not when the constituent stimuli are unrelated. Everyday life is replete with examples of when this ability is important – from holding conversation in noisy environments, detecting distant or faint objects, or appreciating live musical performance, to crossing a busy road, the brain can improve the reliability of the often noisy and incomplete sensory information it has available to make behavioural decisions by integrating inputs from multiple senses. Thus, to survive and prosper in the environment we must appropriately integrate information from each different sense modality. When multimodal cues are generated by the same event, they are usually in close temporal and spatial alignment, and their integration can result in the enhancement of detection, identification, and orientation behaviour to the event.

Experiments using psychophysical methods have repeatedly shown these effects across many combinations of sense modalities. For example, decreased reaction times have been found to auditory-visual stimuli (Gielen et al., 1983, Fort et al., 2002), tactile-visual stimuli (Forster et al., 2002), auditory-tactile stimuli (Murray et al., 2005), and olfactory-visual stimuli (Gottfried, 2003), compared to stimuli presented to either sense alone. Tri-modal audio-visual-haptic stimuli have also shown to be detected faster than any combination of the respective bi-modal stimuli (Hecht et al., 2008).

The neural mechanisms by which multimodal information is integrated have been investigated only relatively recently. Up until the early 1970s, sensory systems were largely studied in isolation, with the view that uni-modal information was processed to a highly elaborated state independently, with integration only occurring in high level poly-sensory or association cortical areas (Felleman and Van Essen, 1991, Rauschecker et al., 1997). Pioneering work in the 1970s and 1980s started to challenge this established view, with several groups finding evidence of neurons in cortical and mid-brain structures that showed evidence of multisensory receptive fields (King and Palmer, 1985, Stein and Arigbede, 1972, Seltzer and Pandya, 1980). The cat superior colliculus (SC) has been used as a model system for MSI, and much of the initial work on this structure is
summarised in an influential book called The Merging of the Senses, by Barry Stein and Alex Meredith (Stein and Meredith, 1993). The book summarises a wealth of neurophysiological data, mostly from the cat SC, and describes the conditions under which multisensory neurons in the SC integrate their inputs. The distinction between neurons that have multisensory receptive fields and those that also integrate their inputs is important, and will be re-visited in Section 2.3.3.1.

2.2.1 Auditory effects on visual perception

The dominance of vision over audition can be seen in many areas of behaviour, even at the level of society and culture as a whole. Consider the statement ‘I’ll believe it when I see it,’ which might be uttered by a speaker when some unlikely event is alleged to have occurred. The statement implies that the speaker will only believe that the event actually occurred if she or he visually ‘sees’ evidence. This bias has carried over into the scientific community, with visual perception often viewed as a dominant modality. As has been noted already, the visual pathway from the eye to the high levels of the brain was thought to be modular and separate from the other modalities. Although some effects of auditory stimuli on visual perception have already been described above, the effects of visual stimuli on auditory perception are far more numerous and well-studied. There are many strong examples of visual-auditory illusions, such as the McGurk effect (McGurk and MacDonald, 1976) and the ventriloquist’s illusion (Bertelson, 1999), in which vision of mouth movements affects the perception of speech sounds and the spatial location of a speaker respectively. By comparison, the list of research into situations in which an auditory stimulus can affect visual perception has historically been much smaller. Recently, however, the realisation that the sensory systems are much more inter-connected than once thought has driven an increased interest in the area.

As well as reaction time and detection threshold advantages, multimodal stimulus presentations can have noticeable perceptual effects. For example, the simultaneous presentation of a tone can improve detection of a dimly flashed light (Frassinetti et al., 2002a, Frassinetti et al., 2002b), enhance the discriminability of
briefly flashed visual patterns (Vroomen and de Gelder, 2000), and increase perceived luminance (Stein et al., 1996). Sound can also affect temporal features of visual perception. For example, the perceived duration of a visual stimulus, as well as the perceived gap between two successive visual stimuli can be affected by the duration of a co-occurring sound (Walker and Scott, 1981). There are also numerous studies describing how the perceived rate of a flickering visual stimulus can be affected by changes in rate of an auditory flutter stimulus (Gebhard and Mowbray, 1959, Recanzone, 2003, Shipley, 1964, Welch et al., 1986), and even temporal order judgements of two visual stimuli can be improved by a sound presented immediately prior and after each visual stimulus (Morein-Zamir et al., 2003, Hairston et al., 2006).

Auditory-visual illusions are examples of even stronger effects of auditory stimuli on visual perception, and will be covered in detail in the next chapter. In the following sections, some possible neural pathways and neural correlates of the processes that underlie MSI are explored, after a brief overview of the uni-sensory processing pathways.

### 2.3 Neural pathways for auditory, visual and auditory-visual processing

The hierarchical structures of the auditory and visual systems have been studied for decades. Although an in-depth review of the auditory and visual pathways is beyond the scope of this thesis, a brief discussion of the major anatomic landmarks is in order. The major structures of the ascending auditory pathway are illustrated in Figure 2-1.

#### 2.3.1 The auditory system

##### 2.3.1.1 Outer, middle and inner ear

The auditory system mediates the sensory experience of hearing. Sound waves travel through the air as wave fronts of compression and rarefaction of air molecules produced by a vibrating object in the environment. Upon reaching an
individual, these pressure waves are reflected towards the ear canal by the shoulders, head, and external parts of the ear, and are transmitted through the ear canal to the eardrum or tympanic membrane, where they cause small displacements of the membrane. These tiny movements, of as little as $10^{-9}$ m (Altschuler et al., 1991), are amplified by a system of three small bones called the middle ear ossicles. The last of these bones acts like a piston, and transmits vibrations into the fluid-filled cochlea. The cochlea is a spiral-shaped structure embedded in the temporal bone, and contains rows of hair cells lined up along the length of the spiral. These hair cells are mounted on the basilar membrane, which has mechanical properties causing it to resonate at different frequencies along its length. The hair cells are thus set in motion at different points along the membrane depending on the frequency of vibration transmitted through the eardrum and ossicles into the cochlea. This organisation of frequency along the receptor surface is called tonotopic mapping, and is maintained at various stages throughout the auditory system into the auditory cortex.

### 2.3.1.2 Brainstem

Although there are nerve tracts descending from the cortex, which pass through all the nuclei of the auditory system to the cochlea (see dashed lines in Figure 2-1), the ascending pathways from the cochlea to the brain will mainly be discussed here (solid lines in Figure 2-1).

---

1 Note that in the visual and somatosensory systems, the topographic relationship between the receptor surface and the arrangement of nerve fibres in the brain is inherently spatial. In the visual and somatosensory systems, the receptor surfaces have a direct spatial relationship with stimuli in the environment – thus stimulus location is automatically coded. In the auditory system however, the natural arrangement of the cochlea maps the frequency of sounds – spatial properties of stimuli are determined partly in the brainstem by comparing timing, loudness, and spectral differences between the signals at each ear.
Figure 2-1. The main ascending (solid lines) and descending (dashed lines) auditory pathways from the left cochlea to the primary auditory cortex. Figure and original caption adapted from (Brugge and Howard, 2003). AC – auditory cortex; MGB – medial geniculate body; IC – inferior colliculus; NLL – nucleus of lateral lemniscus; MNTB – medial nucleus of the trapezoidal body; LSO – lateral superior olivary nucleus; MSO – medial superior olivary nucleus; PO – pre-olivary nucleus; DCN – dorsal cochlear nucleus; VCN – ventral cochlear nucleus; AN – auditory nerve (8th cranial nerve).
Hair cells on the basilar membrane activate primary sensory neurons in the spiral ganglion. Signals from the sensory neurons ascend the 8th cranial nerve and project to the first brainstem nucleus – the cochlear nucleus (CN). The CN is divided into three sub-nuclei – the dorsal (DCN), posterio-ventral (PVCN) and anterio-ventral (AVCN) cochlear nuclei. Key functions of the VCN and DCN are horizontal localisation and the control of hair-cell sensitivity, respectively. From this point on, multiple parallel pathways ascend the brainstem to the auditory cortex. Axons leaving the VCN project to the superior olivary complex (SOC), and then via the lateral lemniscus to the inferior colliculus (IC) in the midbrain. The pathways from each ear are mainly separate until the SOC, at which point they converge, allowing further sound source localisation. All the parallel pathways in the brainstem converge at the IC, which is both tonotopically as well as spatially organised. While the VCN is concerned with horizontal localisation, the DCN is involved with vertical localisation. Axons from the DCN project via the lateral lemniscus to the contralateral IC, skipping the SOC.

2.3.1.3 Midbrain

The inferior and superior colliculi (IC and SC) together form the highest level of the midbrain, and both are major integrating structures, receiving input from many sensory systems. The SC will be covered in more detail in section 2.3.3.1. The IC is a main link between the auditory brainstem nuclei and the rest of the brain. It receives input from many other sensory and cortical regions, and projects just as widely – up to higher cortical regions as well as back down to many of the ipsi-lateral auditory brainstem nuclei. From the ascending auditory system, the IC receives input from three main pathways: 1) from the ipsi- and contra-lateral superior olivary complex (lateral and medial superior olivary complex, LSO and MSO in Figure 2-1), 2) a direct pathway from the contra-lateral CN, and 3) the ipsi- and contra-lateral lateral lemniscus. The IC is thought to be involved with determining relatively high-level features of sound, such as pitch, onset envelopes, and intensity. The IC projects to the auditory forebrain through the brachium of the IC, and to the SC through the lateral nucleus of the IC.
2.3.1.4 Auditory forebrain - thalamus and cortex

The medial geniculate body (MGB) of the thalamus and multiple areas of cortex in the temporal lobe form the auditory forebrain. The MGB receives input from the IC and projects to the auditory cortex (AC), located in the lateral sulcus on the superior surface of the temporal lobe (see Figure 2-2). The ventral division of the MGB is sharply tuned for frequency, and receives input from the central part of the IC, which also has a strong frequency organisation. The ventral division projects to the core regions of AC. Dorsal MGB neurons, with wider receptive fields, project to belt areas of AC.

Figure 2-2. Diagram showing the location of the primary and secondary auditory (A1, A2) and visual (V1, V2) cortex in the human brain. V1 and V2 are located bilaterally on the occipital poles. Although only V1 and V2 are highlighted, many areas within the dotted line are known to be associated with visual processing). A1 and A2 are only partly shown on this view, as they are located partly within the superior temporal sulcus, on the superior surface. Brain outline from (Rosa, 2003), functional regions added after consultation with the Brodmann interactive atlas (Bernal and Perdomo).

The AC has a hierarchical organisation with three main levels – the core area (A1), and the belt and parabelt areas surrounding the core (A2 and A3). The AC is distributed partly on the exposed surface of the superior temporal gyrus
(STG) and partly buried within the Sylvian fissure (see Figure 2-2). The core area contains the primary auditory cortex (PAC or A1) and receives tightly frequency-tuned input from the ventral MGB. The core area has a columnar organisation, with columns of neurons sharing similar characteristic frequencies (CF) and binaural response properties (Middlebrooks, 2009). The belt areas (A2) receive feed-forward projections from the core as well as direct projections from the dorsal MGB. A third surrounding layer, the para-belt region (A3), receives feed-forward projections from the belt area (but not the core), and itself projects to other cortical areas. It is not known to be highly involved in multi-sensory processes so will not be discussed further here.

2.3.2 The visual system

2.3.2.1 Functional anatomy of the visual system

The human visual system, as with the auditory system, has a hybrid hierarchical/parallel structure, with increasing elaboration of information as it ascends through sub-cortical nuclei to the primary cortical areas and beyond. As in the auditory system, there is a consistent organisation from receptor surface to cortex. In the visual system this takes the form of retinotopic mapping – the spatial mapping of receptor cells on the retina is preserved throughout the ascending visual system to the cortex. Unlike the auditory system, however, the majority of this hierarchical system consists of a large number of highly interconnected cortical areas, with fewer major sub-cortical nuclei. From the retinal ganglion cells, 90% of optic nerve fibres project via the optic chiasm directly to the lateral geniculate nucleus in the thalamus (LGN), and then on to the primary visual cortex (V1) (Møller, 2003). Most of the remaining 10% of optic nerve fibres project directly to the SC, where they have targets on many neurons with multisensory receptive fields. Section 2.3.3.1 will provide more detail on the structure and function of the SC.
Figure 2-3. The projections from the retina through the optic chiasm, LGN, and optic radiation to the primary visual cortex. Figure adapted from Wiencken-barger & Casagrande (2002).

From the LGN, four main streams form the optic radiation, which projects via the wall of the lateral ventricle to V1 (see Figure 2-3). The retinotopic arrangement persists until this level, such that activation on the cortical surface has a spatial pattern matching the gross features of the original stimulus (Sereno et al., 1995). The visual cortex occupies a large part of the human brain, occupying virtually all of the occipital lobes, and extending into the parietal and temporal lobes (see Figure 2-2, although only V1 and V2 are shown, visual areas extend to the dotted line). The visual cortex is subdivided into many different functional areas, which all are extensively interconnected. The primary visual area, V1, is located bilaterally on the occipital poles, and extends into the calcarine sulcus. V1 maintains the retinotopic mapping inherited from the subcortical structures, and is connected to many further visual areas (Felleman and Van Essen, 1991). The function of these areas is beyond the scope of this review; however it is interesting to point out this major difference in organisation compared with the auditory system, which has far more subcortical involvement.
2.3.3 Auditory-visual integration

After signals are transduced by the sense organs, information is progressively elaborated upon through a system of feed-forward and feed-back connections from each of the ascending sensory pathways as well as other areas of the brain. Two main areas of integration are the SC, and various cortical areas including the anterior ectosylvian sulcus (AES) in the ‘association regions’ of the parietal lobe. There is increasing evidence that neurons in regions previously thought of as uni-sensory, such as V1 and the belt areas of AC also show multi-sensory characteristics (Kayser and Logothetis, 2007, Kayser et al., 2005, Foxe et al., 2002).

2.3.3.1 The Superior Colliuli

The best-known sites of MSI are the SC. The primary function of the SC is in orienting – to direct behaviour to the point in body-related space from which a stimulus has been detected. The SC are multi-layered bilateral mid-brain structures that receive afferents directly from the retina, feedback from various cortical regions (mainly visual cortex), the spinal cord, and the inferior colliculus in the ascending auditory pathway. They send output to the spinal cord, cerebellum, thalamus and visual cortex via the LGN. While cells in the superficial layers (layers 1-3) of the SC are primarily visual, neurons responsive to stimulation of multiple senses have been located in the deep layers (4-7) of the SC in cats (Meredith and Stein, 1983, Peck, 1987), monkeys (Jay and Sparks, 1984) and several rodent species (Wallace et al., 1996). These neurons receive afferents from auditory, visual, and tactile modalities in humans, as well as infrared sensation in snakes (Hartline et al., 1978) and electric field sensation in snakes and sharks, rays and skates (Bodznick, 1990).

The inputs to the SC are topographically arranged, such that a given neuron will have the same receptive field for each of the modalities it is sensitive to. In this way, a neuron receptive to visual stimuli in the far left visual field will also be
sensitive to auditory inputs from the left side. In addition, the receptive fields stay constant no matter the source of the afferent – whether it is directly from the sensory epithelium, a sub-cortical structure, or sensory cortex. The same neuron may thus have a receptive field map made up of both cortical and sub-cortical afferents, all with consistent spatial organisation. The receptive field maps are slightly overlapping, such that stimulation from the same spatial location in each modality activates the same region of neurons (King and Palmer, 1985, Meredith and Stein, 1996). These sensory maps are also aligned with motor representations, allowing stimulation from a given location in any sense to initiate the same motor or orienting response (Stein and Meredith, 1993).

As well as providing a common representation of external space (or perhaps as a consequence), the SC is known to be involved in the functional integration of sensory information, with some multisensory neurons also conditionally responding in a non-linear manner when receiving stimulation from multiple senses. Figure 2-4 shows the non-linear super-additive response properties of a multisensory neuron in the monkey SC. The conditions under which these individual neurons respond to multisensory stimuli have been extensively studied, particularly in cats and monkeys, and these conditions have been summarised by three “rules” of multisensory integration (Stein and Meredith, 1993). These three rules, which describe the responses of single neurons in the SC in relation to their inputs, have also generally been found to describe the relationship between behavioural outcomes (such as the enhancement of reaction times or detection rates described in Section 2.2) in relation to external stimuli in humans (Stein et al., 2009a).

The “spatial rule” states that neural responses tend to be enhanced when the stimuli in different senses come from the same location in space, the “temporal rule” when the stimuli occur at the same time, and the “inverse effectiveness” rule when at least one of the stimuli by itself is only weakly effective in exciting the given neuron.
Figure 2-4. An example of non-linear response enhancement for a multisensory neuron in the monkey SC. The top panel shows the visual and auditory spatial receptive fields for the neuron. The middle panel shows raster plots and histograms of the spiking activity of the neuron. The bar graph to the right shows the mean spikes per trial, and indicates that the number of responses in the visual + auditory condition was significantly higher than in either the auditory or the visual condition, and also higher than the sum of the auditory and visual conditions (indicated by the dashed line). Figure with original caption from Wallace, Wilkinson & Stein (1996).

2.3.3.1.1 Three rules describing neural responsiveness in the SC

The spatial rule is thought to be a direct consequence of the manner in which the maps of receptive space in the SC are closely aligned across modalities. As long as signals from two (or three) modalities are within the space defined by
the overlapping receptive fields, the response of the neuron will be enhanced, and
the stimuli will be registered as having a common origin in the environment (see
Figure 2-4). If one stimulus falls outside the neuron’s receptive field, the neuron’s
response will not be enhanced. Each excitatory receptive field also has a
surrounding inhibitory region – if one stimulus falls inside this inhibitory region,
the neuron’s response will be reduced rather than enhanced (Meredith and Stein,
1996).

A complicating factor in maintaining aligned receptive fields between either
auditory or somatosensory and visual sensation is the ability to move the eyes – as
the eyes move, retinotopic mapping alone cannot provide an accurate indication of
the stimulus location in external space. In order to overcome this problem,
auditory and somatosensory receptive fields are linked to the position of the eyes.
The issue of maintaining spatial register between visual, auditory and
somatosensory references frames is still under investigation and is beyond the
scope of this review, however there are several reviews available on the topic
(Groh & D. L. Sparks, 1996; Hartline, Vimal, King, Kurylo, & Northmore, 1995;
Jay & David L. Sparks, 1984; Peck, Baro, & Warder, 1995).

The temporal rule has been similarly explained in terms of neural properties
of the SC, and is thought to be a consequence of the fact that although neural
spikes only last several milliseconds, the excitatory and inhibitory currents within
neurons may persist for several hundred milliseconds (Meredith et al., 1987). This
slight flexibility in the timing of signals arriving at the SC takes into account
differences in the conduction of different stimulus energies through the
environment (light travels faster than sound in air), as well as differences in the
processing speed of the difference sensory receptors and processing pathways
themselves. Generally, the magnitude of the integrated response is largest when
the peak activity in each sense overlaps.

One of the most common observations concerning multisensory processes
is that the largest benefit often occurs when one or both of constituent stimuli are
in themselves degraded or near perceptual thresholds. This makes intuitive sense –
in the environment, uni-modal stimuli that are easily perceived will be easily detected and identified with a high degree of certainty. In noisy environments however, or when a stimulus is weak, it makes biological sense that confirmatory information from other senses should be utilised when it is available. Likewise in the SC, the biggest neural response enhancement occurs when the uni-modal constituents of a multimodal stimulus elicit only a small response by themselves (Stanford et al., 2005, Stanford and Stein, 2007).

2.3.3.2 The role of cortical input to the SC

Inputs from multiple areas of cortex are also essential for MSI in SC neurons. In cats, these descending inputs come mainly from the parietal association cortex: the AES and an adjacent area, the rostral aspect of the lateral suprasylvian sulcus (rLS) (Jiang et al., 2001, Jiang et al., 2002). When these inputs are blocked by reversible cooling, the target neurons in the SC still respond to uni-modal stimuli within their receptive fields, but no longer show the capacity to functionally integrate the inputs (Alvarado et al., 2007). The AES has a number of subdivisions which can be defined based on the modality of the target neurons in the SC. A target neuron in the SC that receives auditory-visual input for example, needs to receive convergent input from the auditory and visual subdivisions of the AES in order to integrate its inputs (Wallace et al., 1992).

As well as reducing the ability of SC neurons to integrate their inputs, blocking parietal input to the SC compromises behavioural responses. The enhanced ability to orient towards and localise multimodal compared to uni-modal stimuli is reduced, but the ability to respond to each uni-modal component is not (Wilkinson et al., 1996). The mechanism by which these cortical inputs enable integration in the SC is currently unknown. The functional roles of cortical input are thought to include a mechanism for learning and plasticity – these parietal cortex inputs may be the main mechanism by which environmental factors influence the development and adaptation of appropriate integration function in the SC (Stein et al., 2009b). These issues are further discussed in the introductions.
to the three experimental chapters (Chapters 5, 6 and 7) and in the general discussion (Chapter 8).

2.3.3.3 Integration in cortical areas

Cortical neurons, including but not limited to those in the parietal association cortex that project to the SC, have also been found to have multisensory receptive fields, and sometimes display super-additive behaviour themselves. The conditions under which these cortical neurons display integrative responses are also more complex than those in the SC. Rather than requiring spatial and temporal congruence, cortical neurons often require semantic congruency. For example, activity measured by fMRI in superior temporal sulcus (STS) only shows enhanced responses to multimodal stimuli when the video and audio is semantically matched (Beauchamp et al., 2004).

Compared with the SC however, super-additivity in cortical areas is rare (Kayser and Logothetis, 2007). Whereas multisensory neurons in the SC can have their firing rates altered by an order of magnitude in response to multisensory stimuli, cortical MSI is far more likely to take the form of small alterations to firing rates. Cross-modal influences in cortex are thus more of a modulatory nature. Secondly, the distribution of multisensory neurons is different. In the SC (and higher temporal and frontal association cortices), only a small subset of neurons integrate inputs from multiple senses, but those that do can show very strong response enhancement or depression. In comparison, multisensory neurons are widespread within the sensory cortices, but show only small modulations to multisensory stimuli (Meredith et al., 2009).

2.3.3.4 Cortico-cortical projections

As well as receiving feed-forward connections from the subcortical modality-specific pathways and the SC, the auditory and visual primary cortical areas have recently been found to have direct monosynaptic connections with each other, as well as with regions of the parietal lobes (Falchier et al., 2002, Rockland and Ojima, 2003, Clavagnier et al., 2004). These projections arise from the parietal
lobes, and the belt and parabelt areas of the auditory cortex, and have targets in primary and secondary visual cortex, mostly in regions that subserve the peripheral and lower visual fields.

Figure 2-5. The proportion of neurons in the visual cortex (at eccentricities from 0-20 degrees) with direct connections to auditory cortex (light bars) and the polysensory area of the superior temporal lobe (STP – black bars). Note that the proportion of neurons with auditory connections increases with increasing eccentricity, and that the overall proportions are less than that received from STP. Figure and original caption from Falchier et al (2002).

Figure 2-5, from Falchier et al (2002), shows the proportion of neurons in the primary visual cortex that have a direct connection to the auditory cortex (light bars) compared to inputs from the superior temporal lobe (STP, black bars). The STP is another polysensory area with known connections to the visual system. Overall, it can be seen that the connections with auditory areas are far sparser than with STP. In foveal areas (0-2° eccentricity), there are almost no connections with
auditory cortex. However, as the eccentricity increases up to 20°, well into the visual periphery, up to 12% of primary visual cortex neurons were found to have a direct, monosynaptic connection with the auditory cortex. These findings tie in with behavioural and ERP data in humans indicating that modality-specific judgements (such as contrast for vision) are sensitive to the spatial congruence of auditory stimuli (McDonald et al., 2003, Bolognini et al., 2005a, Frassinetti et al., 2002a, Frassinetti et al., 2002b). In addition, the earliest (~50 ms) multisensory effects generally occur irrespective of the stimulus location (Murray, et al., 2005), but are sensitive to temporal congruence (Senkowski et al., 2007).

In humans, similar connections have been shown using fMRI (Eckert et al., 2008). During resting state scanning, the authors found that activity in medial Heschl’s gyrus was strongly coupled to the visual cortex, along the anterior banks of the calcarine fissure, an area known to represent the peripheral visual field. When participants were subsequently asked to perform a simple visual task, overall coupling between primary auditory and visual areas reduced, but the specific coupling from auditory cortex to the anterior calcarine cortex was not affected. Cate et al (2009) have also shown that auditory activations in the peripheral visual cortex are dependent on auditory attention – only attended sounds (outside the visual field) activated peripheral visual cortex. Most recently, Beer, Plank and Greenlee (2011) used resting-state diffusion tensor imaging to show that seed areas in Heschle’s gyrus (containing the primary auditory cortex) projected strongly to the ipsilateral temporal, parietal, and occipital cortex. In the occipital lobe, large clusters of tracts seeded in Heschle’s gyrus projected to the ventral occipital cortex, with smaller clusters found targeting the anterior part of the calcarine sulcus and the occipital pole. However, when the tracts were seeded in the planum temporale, containing the secondary auditory cortex, they were less likely to project to the occipital areas.
Figure 2-6. Diagram showing the location of the primary and secondary auditory (A1, A2) and visual (V1, V2) cortex in the human brain. V1 and V2 are located bilaterally on the occipital poles. A1 and A2 are only partly shown on this view, as they are located partly within the superior temporal sulcus, on the superior surface. Grey areas from light to dark shades show progressively higher levels of the visual processing stream.  

A) ‘traditional’ view of the integration of auditory and visual information, where all sensory information is progressively elaborated throughout uni-sensory pathways before being integrated in higher ‘association cortex.’  

B) Strictly cortico-cortical connections only.  

C) Emerging view where, in addition to the flow
of information to the higher areas of cortex (grey lines), there are also low-level connections directly between primary sensory cortices. The thick line indicates that evidence has been found for more numerous connections between the primary sensory areas than between secondary areas. Reciprocal connections between A1-A2 and V1-V2 are presumed to exist in both models but are excluded for clarity. * A1 -> V1: (Rockland and Ojima, 2003, Falchier et al., 2002, Clavagnier et al., 2004). ** A1-V2 (Rockland and Ojima, 2003) *** V2 -> A1: (Budinger et al., 2006).

The results from both these studies in humans are thus consistent with the previously mentioned tracer studies in primates (Clavagnier et al., 2004, Falchier et al., 2002, Rockland and Ojima, 2003) that have found direct axonal connections between the parietal lobes, auditory cortex and both the primary and secondary visual cortex. Such connections are relevant to the study of rapid multi-sensory interactions, as many of the behavioural consequences of multi-sensory interactions, especially those associated with illusions, occur on a timescale too rapid for the traditional hierarchical model to support. Figure 2-6 shows two possible models supporting multi-sensory interactions. Panel A shows the traditional model, where parallel and separate uni-sensory processing streams are not merged until late, high-level ‘association cortex.’ Panel B shows a model where only cortico-cortical connections exist. Panel C shows the newly-emerging view of how multisensory processing might be substantiated in the brain, by interactions directly between primary and secondary sensory cortices, with feedback as well as feed-forward connections from association regions influencing MSI at both a sub-cortical and cortical level.

2.4 Electrophysiological correlates of multisensory integration

As has been shown, there is anatomical evidence suggesting that integration of auditory and visual information can occur at the level of the primary visual cortex, via both feed-forward and feedback type connections. Multimodal stimuli
can affect both ERPs (Giard and Peronnet, 1999, Molholm et al., 2002, Shams et al., 2005a, Brefczynski-Lewis et al., 2009) and gamma band responses (Mishra et al., 2007, Senkowski et al., 2006, Senkowski et al., 2008) during early sensory processing, at latencies of 30-60 ms, over occipital regions of the brain. However it is still unclear whether both auditory and visual inputs to visual cortex can actually arrive within this timeframe. A brief description of the neural propagation and timing of these systems may thus be instructive, as the temporal constraints will necessarily define the time limits over which auditory information can have an effect on visual processing. There are ‘late’ effects of multi-modal stimuli on uni-sensory processing (see Section 3.6 for an example), but the focus here is on the shortest times in which these effects can occur.

After the ascending auditory pathways, local field potentials have been recorded using subdural electrode arrays on human primary auditory cortex from between 15-20 ms following the presentation of simple stimuli such as clicks and short noise bursts (Howard et al., 2000, Brugge et al., 2003, Liegeois-Chauvel et al., 1994). Post-synaptic potentials have also been recorded from parietal and frontal cortex around 50-100 ms following a variety of auditory stimuli – from simple clicks and pips (Giard et al., 2000, Spierer et al., 2007), to more complex sounds such as speech (Besle et al., 2008). This spread of activation is consistent with anatomical studies in primates (Kaas and Hackett, 2000, Kaas et al., 2008), which suggest that the primary auditory cortex is separated from middle temporal and frontal regions by only 2 or 4 synapses, respectively.

Recent research into the development of appropriate stimulus coding schemes for retinal implants has shown that cortical activation in V1 occurs in two distinct phases after an electrical stimulus is delivered to the cat retina. There is an initial transient response around 10 ms, followed by a slower component from 40-55 ms (Elfar et al., 2009, Shivdasani et al., 2010). Whereas the initial transient response is elicited by widespread retinal locations, the later response originates from a more focussed retinal area. Similarly in monkeys (Chen et al., 2007), initial excitation of neurons in V1 was found to occur at around 26-32 ms post-stimulus. In humans, the C1 visual ERP, thought to index the first activation of V1 and V2
by retinotopic afferents, has an onset time of around 45-60 ms, although the exact time of the first visual responses in V1 is difficult to measure and interpret (Foxe et al., 2008). However, a rough 3/5ths rule (monkey latencies are around 3/5ths of the human equivalent) can be applied to convert conduction times in the monkey to human timescales (Schroeder and Foxe, 2004). Applying this rule to the direct measurements from V1 in monkey thus gives an estimated first visual response time in the visual cortex of humans of around 42-51 ms, corresponding relatively well with the C1 scalp-measured ERP. Given that modulations of visual ERPs by multisensory stimuli have been found to occur in the 30-60 ms timeframe, there is little time for extensive feedback systems from high-level cortical areas, and the interaction is likely to occur at a relatively low level, perhaps by the direct cortico-cortical connections detailed previously.

### 2.5 Development

Although there has been a great deal of research investigating the development of MSI in animal models (Stein et al., 2009b, Polley et al., 2009, Wallace and Stein, 2007), and even in human infants (Lewkowicz and Turkewitz, 1980, Scheier et al., 2003a, Neil et al., 2006, Pons et al., 2009), there has been less focus on school-age children. As we have seen, effective MSI can improve many perceptual outcomes, especially when the signal to noise ratio is low. Schools and other learning situations are often environments where signals of interest are presented in a background of noise or other competing signals both social and educational. It is thus of great practical interest to start to understand how the effective integration of multimodal signals develops during childhood. Chapter 7 of this thesis consists of a published study (Vongpaisal et al., 2006) with a short introduction covering these issues, however for completeness a brief review of the literature covering multisensory abilities in children is repeated and expanded upon here.
2.5.1 Matching and integration

As was discussed in Section 2.1, there has been some degree of confusion of terminology in the literature, especially where research from multiple disciples has been combined. This is especially apparent in the multisensory development literature, where results from neurophysiological investigations of the development of multisensory structures in animal models, using single-cell recordings of multisensory neurons, must be combined with results from neuropsychological investigations of human infants and children. The main confusion has been between the idea of multisensory integration abilities, and cross-modal matching abilities. The ability to match or compare information from multiple sensory signals appears quite early in human development. For example, infants can match the intensities of auditory and visual signals (Lewkowicz and Turkewitz, 1980), discriminate between redundantly-specified auditory-visual tempos (Bahrick et al., 2002), and at only 1-3 days post-natal are able to match non-human primate vocalisations to the appropriate visual gesture (Lewkowicz et al., 2010). However, as with many uni-sensory abilities (such as speech perception (Kuhl et al., 2003), face perception (Pascalis et al., 2002), and music perception (Hannon and Trehub, 2005), some multisensory abilities present at birth or soon after show evidence of narrowing or tuning with post-natal development. For example, the ability to respond to an audio-visual non-human primate phonetic contrast has been shown to narrow between 6 and 11 months of age (Pons et al., 2009). This finding and many others (for review see Lewkowicz and Ghazanfar, 2009) suggest that post-natal experience with the appropriate stimulus-response contingencies is required in order to tune the matching processes to best suit the child’s environment.

Multisensory abilities such as intensity matching (Lewkowicz and Turkewitz, 1980), and the detection of temporal equivalence in the form of rhythm (Bahrick and Lickliter, 2004, Bahrick and Lickliter, 2000) or speech onsets and offsets (Lewkowicz, 2010, Lewkowicz et al., 2010), can appear early in infancy. Although these abilities appear to be present at birth (and are often subsequently refined or narrowed with post-natal experience), other multisensory phenomena emerge later in life. For instance, in the stream-bounce illusion (Sekuler et al.,
1997), the presentation of a brief click as two moving balls intersect on a video display biases the perception towards bouncing rather than streaming in adults. In infants, however, this illusion is not perceived until at least 8 months of age (Scheier et al., 2003a). In the McGurk illusion, incongruent visual stimuli have less of an affect on final phoneme perception in both 3-5 and 7-8 year old children than in adults (McGurk and MacDonald, 1976, Massaro, 1984, Massaro et al., 1986).

In both these cases, information from one sense alters or enhances perception in the other, requiring a more integrative process than tasks requiring the transfer or matching of information across the senses. Gori, Del Viva, Sandini, and Burr (2008) have likewise shown that judgements of size and orientation do not benefit from the ability to optimally integrate visual and haptic information until between 8 and 10 years of age, and Nardini, Jones, Bedford, and Braddick (2008) have shown that children up to 8 years of age do not integrate self-generated motion cues with visual cues in an object navigation task. Barutchu, Danaher, Crewther, Innes-Brown, Shivdasani and Paolini (2010), using a simple auditory-visual detection task, also found that reaction times to auditory-visual stimuli do not show mature super-additive enhancement in children as old as 11 years of age.

In summary, although infants can match information across the senses at a very young age, the facilitation of perceptual judgements and reaction times by the integration of multimodal stimuli seems to develop slowly throughout childhood. As mentioned in Section 2.3.3.2, input to the SC from cortical areas such as the AES in the parietal association cortex, which are much slower to develop than brainstem areas, may be responsible for the fine-tuning of integrative abilities, guided by development and experience.

2.6 Conclusion

There is still much debate over the possible neural pathways that allow the modulation of visual processing and perception by sounds. ERP studies have shown that sounds can modulate early visual ERPs generated in the occipital
cortex (Beer and Watanabe, 2009, Beer and Röder, 2004, Beer and Röder, 2005, Eimer and Driver, 2000, McDonald et al., 2003, McDonald et al., 2005), early as 40 ms post-stimulus (Molholm et al., 2004, Molholm et al., 2002, Foxe and Schroeder, 2005). However, traditional models, where multisensory integration occurs only in poly-sensory or associative cortex, fail to explain these effects, as the response latencies of higher-level cortex are too slow – feedback loops from high level cortex are thus unlikely to account for the rapid effects of sounds on visual processing. As has been seen, direct cortico-cortical connections have recently been found in primates using tracer techniques (Clavagnier et al., 2004, Falchier et al., 2002, Smiley and Falchier, 2009), and in humans using resting state fMRI and diffusion tensor imaging (Eckert et al., 2008, Beer et al., 2011). These connections would allow rapid transfer of information between the senses at a very early stage of processing. This rapid transfer could explain the very early ERP effects, and behavioural effects such as rapidly-evolving auditory-visual illusions (to be reviewed in the next chapter). In primates, these connections have also been found to have spatial properties that correlate well with behavioural accounts of auditory-visual effects. While the connections from auditory cortex to visual cortex have been shown to proceed from mostly auditory core areas to peripheral areas of the visual field, sounds from a wide spatial extent have been shown to affect visual processing only at very limited areas of the visual field (Bolognini et al., 2005a, Frassinetti et al., 2002a, Frassinetti et al., 2002b).

In this introductory chapter, the basic mechanisms of perception in vision and audition were briefly reviewed. Sub-cortical and cortical mechanisms underlying multisensory integration were covered, and the development of multisensory abilities and neural mechanisms in children were discussed. In the later experimental chapters, an audio-visual perceptual illusion will be used as an index of MSI in the brain. In the chapter to follow, a brief history of this illusion will be provided.
3 The “Sound-induced Flash Illusion”
Chapter 2 introduced anatomical models of MSI – mainly derived from studies of the SC and other sub-cortical structures in animals – as well as the functional significance of MSI in humans, and behavioural and psychophysiological studies of the phenomenon in humans. The literature has converged in recent years towards the idea that in addition to sub-cortical MSI, some kinds of integration between the senses can also occur within and between areas of primary sensory cortex. As reviewed in Chapter 2, there are many different paradigms that have been employed to assess MSI in humans. One of these is the so-called ‘sound-induced flash illusion’ or ‘flash-beep illusion’ (Shams et al., 2000), in which two distinct flashes are perceived when a single light flash is presented along with two beeps. The flash-beep illusion is well suited for investigating the effect of one sensory modality on another, as it is possible to tell, on a trial-by-trial basis, whether or not integration between audition and vision has caused the perception of an additional flash. After a brief comment on perceptual illusions, this chapter aims to chronologically review a selected portion of the increasing number of existing studies employing the flash-beep illusion, so that the results using the paradigm reported in this thesis may be placed in context. Although different authors use different terminology for the various combinations of flashes and beeps that can be used, the notation ‘XFYB’ will be used through this thesis, indicating X flashes and Y beeps. ‘1F2B’ thus refers to a single flash paired with two beeps.

### 3.1 Perceptual Illusions

Our experience of the world around us is built up by the brain from remarkably incomplete signals transduced by specialised receptors in our sensory epithelia. For the most part our experience of the world, or perception, appears to correlate well with physical reality. However this is not always the case. Perceptual illusions occur whenever there is dissociation between perceptual experience and physical reality. The study of illusions thus provides a window into the operations the brain is performing in building our perceived reality.
Illusions have been studied in both visual and auditory neuroscience. One of the first visual illusions observed was the phenomenon of Mach bands (reviewed in (Purves et al., 2004)). Ernst Mach documented this illusion in 1865—it describes the presence of illusory bands of brightness and darkness at the edges of a luminance gradient. Mach suggested that the illusion occurred due to inhibition between neighbouring points on the retina. This idea, in combination with Sherrington’s burgeoning ‘neuron doctrine’ (for a historical review see Burke, 2007), led to the idea, now fairly well accepted in the visual system, that neurons excite or inhibit their neighbours depending on their patterns of interconnectivity.

Of particular relevance to the study of perception are a class of illusions related to multi-stable stimuli. These are perceptually ambiguous stimuli that have two or more associated perceptual states. In the visual domain, two examples are the Necker cube and Rubin’s vase illusion. In these examples, although the physical stimulus remains identical, the perceptual sensation can alternate between two possible interpretations. Since nothing about the physical stimulus is changing, one can reasonably assume that processes in the brain must be responsible for these effects. More recently, studies on binocular rivalry (presenting different images to each eye) have begun resolve the question of whether such rivalry is caused by a ‘bottom-up’ competition between the two eyes, or between two competing ‘top-down’ cortical explanations for the inputs (Logothetis et al., 1996).

A large proportion of the research into multisensory integration makes use of paradigms where behavioural or physiological responses to multi-modal stimuli are compared to the same responses to either and/or both of the uni-modal stimuli presented alone. For the most part, the uni-modal component stimuli are congruent—that is, there is a natural spatial, temporal or semantic equivalence between the features of each uni-modal part of the stimulus. For instance, in Barutchu et al (Barutchu et al., 2010) a brief noiseburst (A) was paired with a single flash of a checkerboard stimulus (V). Both occurred in the same location and at the same time, and participants were simply asked to respond with a button press as quickly and accurately as possible whenever they heard the sound, saw the flash, or saw/heard both stimuli at once. When the two stimuli were presented together
participants responded faster than could be explained by the summation of response time probabilities from each stimulus when they were presented alone. In this type of paradigm, it is thought that the reduction in reaction time is thus due to integration of the component parts of the stimulus in the brain, rather than simply a response driven by whichever part of the stimulus can facilitate the fastest reaction time.

This experiment, and many others of a similar nature, present the unimodal component parts of multi-modal stimuli much as they might occur in the environment, with the auditory and visual components of the multi-modal stimulus (AV, in this case), produced at the same location in space, occurring at the same time, and matching in as many other features as possible. Behavioural or neurophysiological measures can then be compared between uni-modal and multi-modal stimuli, and an inference concerning the perceptual or behavioural advantage that the multi-modal stimulus provides can be drawn depending on any differences between the various conditions.

In the case of multi-modal illusions, however, the experimenter deliberately introduces a mismatch between one or more aspects of the constituent uni-modal parts of the stimulus that are presented. In most cases, a uni-modal task is required, while a parameter defining the mismatch between each modality is varied. Depending on the combination of stimulus and response, an inference can subsequently be made not about the advantage that the multi-modal stimulus provided, but rather about the presence or operation of the multi-modal process itself, given the reported alteration in perception that the mismatch may have induced.

In the case of the flash-beep illusion, a transient audio-visual stimulus is used, most often consisting of a brief beep or noise-burst paired with a brief presentation of a checker-board or simple disk stimulus. Participants are asked to count the number of flashes they perceive, while various combinations of flashes and beeps are presented. When the number of flashes and beeps presented is equal, participants have no trouble counting the flashes (at least up to 2-3
presentations). However, when more beeps than flashes are presented (1 flash and 2 beeps, for instance), participants often report the perception of more flashes than are really presented. Various experiments, to be reviewed in the following sections, have shown that participants do appear to actually perceive more flashes than are present, rather than have their response criterion changed for instance. Thus an illusion has occurred, and it can be inferred that some process in the brain has caused the integration of features of the auditory stimulus (the number of stimuli presented) with the visual perception of the flashes. The illusion is somewhat similar to the previously-described multi-stable stimuli, in that the same physical stimulus can elicit two distinct percepts.

Illusion tasks are thus very powerful – if conducted carefully, it is possible to determine on a trial-by-trial basis whether multisensory processes are active. However, the circumstances of this integrative process are quite different to those involved in detection and similar tasks, as in the case of illusion tasks, the resulting percept is of course an illusion, or a deviation from physical reality. This is the price paid in return for the highly efficient neural systems that provide a seamless sensory experience in most circumstances. The multisensory process behind the illusion could thus be considered inefficient or inappropriate in terms of the overall goal of perception, which is presumably to render salient aspects of the environment into consciousness as accurately as possible. Nevertheless, the flash-beep illusion has proved to be a reliable window onto multisensory processes in the brain.

Table 3-1 summarises important aspects of the methods and results from many of the studies reviewed below and will be referred to throughout.
Table 3-1. Summary of flash-beep illusion studies. Basic stimulus parameters and results.

<table>
<thead>
<tr>
<th>Reference</th>
<th>N</th>
<th>Display type and refresh rate</th>
<th>Fusion?</th>
<th>A-V delay (ms)</th>
<th>V (ms)</th>
<th>V SOA (ms)</th>
<th>A (ms)</th>
<th>A SOA (ms)</th>
<th>Flash eccentricity (degrees and direction from fixation)</th>
<th>1F0B (% correct or avg resp)</th>
<th>1F1B (% correct or avg resp)</th>
<th>1F2B (% correct or avg resp)</th>
<th>2F0B (% correct or avg resp)</th>
<th>2F1B (% correct or avg resp)</th>
<th>2F2B (% correct or avg resp)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Shams et al (2002)</td>
<td>8</td>
<td>Disp: CRT Ref: unknown</td>
<td>No</td>
<td>23</td>
<td>17</td>
<td>67</td>
<td>7</td>
<td>64</td>
<td>5 down</td>
<td>1.1</td>
<td>1.1</td>
<td>2.2</td>
<td>2.6</td>
<td>2.5</td>
<td>2.2</td>
</tr>
<tr>
<td>Shams et al (2001)</td>
<td>13</td>
<td>Disp: CRT Ref: unknown</td>
<td>N/A</td>
<td>14</td>
<td>14</td>
<td>67</td>
<td>5</td>
<td>57</td>
<td>8 below</td>
<td>-</td>
<td>-</td>
<td>21%</td>
<td>-</td>
<td>-</td>
<td>92%</td>
</tr>
<tr>
<td>Bhattacharya et al (2002)</td>
<td>16</td>
<td>Disp: CRT Ref: unknown</td>
<td>N/A</td>
<td>14</td>
<td>14</td>
<td>67</td>
<td>5</td>
<td>57</td>
<td>8 below</td>
<td>-</td>
<td>-</td>
<td>25%</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Andersen et al (2004)</td>
<td>10</td>
<td>Disp: CRT Ref: unknown</td>
<td>Yes</td>
<td>0</td>
<td>16.7</td>
<td>67</td>
<td>7</td>
<td>67</td>
<td>5 down</td>
<td>80%</td>
<td>92%</td>
<td>40%</td>
<td>70%</td>
<td>40%</td>
<td>80%</td>
</tr>
<tr>
<td>Watkins et al (2006)</td>
<td>17</td>
<td>Disp: LCD projector Ref: 60 Hz</td>
<td>No</td>
<td>0</td>
<td>17</td>
<td>56</td>
<td>10</td>
<td>56</td>
<td>8.5-10 annulus</td>
<td>85%</td>
<td>92%</td>
<td>65%</td>
<td>91%</td>
<td>88%</td>
<td>95%</td>
</tr>
<tr>
<td>Watkins et al (2007)</td>
<td>14</td>
<td>Disp: LCD projector Ref: 60 Hz</td>
<td>Yes</td>
<td>0</td>
<td>17</td>
<td>46</td>
<td>10</td>
<td>46</td>
<td>8.5-10 annulus</td>
<td>-</td>
<td>94%</td>
<td>34%</td>
<td>-</td>
<td>42%</td>
<td>94%</td>
</tr>
<tr>
<td>Meylan &amp; Murray (2007)</td>
<td>8</td>
<td>Disp: unknown Ref: 75 Hz</td>
<td>No</td>
<td>0</td>
<td>13</td>
<td>65</td>
<td>13</td>
<td>65</td>
<td>6.3 below</td>
<td>92%</td>
<td>91%</td>
<td>47%</td>
<td>95%</td>
<td>89%</td>
<td>98%</td>
</tr>
<tr>
<td>Courtney et al (2007)</td>
<td>18</td>
<td>Disp: CRT Ref: unknown</td>
<td>N/A</td>
<td>0</td>
<td>17</td>
<td>NA</td>
<td>7</td>
<td>60</td>
<td>5 below</td>
<td>-</td>
<td>1.0</td>
<td>2.5</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>McCormick &amp; Mamassian (2008)</td>
<td>8</td>
<td>Disp: CRT Ref: 75 Hz</td>
<td>N/A</td>
<td>0</td>
<td>26</td>
<td>78</td>
<td>26</td>
<td>78</td>
<td>10 left and right</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
</tr>
<tr>
<td>Mishra et al (2007, 2008)</td>
<td>34</td>
<td>Disp: Red LED</td>
<td>Yes</td>
<td>10</td>
<td>5</td>
<td>70</td>
<td>10</td>
<td>70</td>
<td>20 left</td>
<td>87%</td>
<td>91%</td>
<td>63%</td>
<td>67%</td>
<td>56%</td>
<td>87%</td>
</tr>
<tr>
<td>Study</td>
<td>N</td>
<td>Display Type</td>
<td>No. of Participants</td>
<td>V SOA (ms)</td>
<td>A SOA (ms)</td>
<td>A (ms)</td>
<td>V (ms)</td>
<td>A-V Delay</td>
<td>Fusion?</td>
<td>Flash Eccentricity</td>
<td>1F0B (% correct or avg resp)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>------------------------------</td>
<td>----</td>
<td>--------------</td>
<td>---------------------</td>
<td>------------</td>
<td>------------</td>
<td>--------</td>
<td>--------</td>
<td>-----------</td>
<td>---------</td>
<td>-------------------</td>
<td>----------------------------</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Tremblay et al (2007)</td>
<td>38</td>
<td>CRT</td>
<td>No</td>
<td>7</td>
<td>5 below</td>
<td>-</td>
<td>-</td>
<td>19, 28, 30%</td>
<td></td>
<td>-</td>
<td>45, 41, 65%</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Innes-Brown &amp; Crewther (2009)</td>
<td>9</td>
<td>CRT</td>
<td>No</td>
<td>23</td>
<td>57</td>
<td>8</td>
<td>67</td>
<td>80%</td>
<td></td>
<td>-</td>
<td>79%</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Innes-Brown et al (2011)</td>
<td>22</td>
<td>CRT</td>
<td>Yes</td>
<td>0</td>
<td>75</td>
<td>8</td>
<td>75</td>
<td>7.5 below</td>
<td>93, 71%</td>
<td>-</td>
<td>95, 83%, 49, 14%, 14, 15%</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Notes: Summary of results from several key studies using the flash-beep illusion paradigm. Key: ‘N’ – number of participants; ‘Display type (Disp) and refresh rate (Ref)’; ‘Fusion?’ – did the authors find evidence of the ‘fusion’ illusion as well as the ‘fission’ illusion; ‘A-V delay’ – the delay from the auditory to visual stimuli, in ms if presented simultaneously; ‘V (ms)’ – duration of visual flash stimulus; ‘V SOA (ms)’ – time between onsets of visual flash stimuli; ‘A (ms)’ – duration of auditory beep stimulus; ‘A SOA (ms)’ – time between onsets of auditory beep stimuli; ‘Flash eccentricity (degrees and direction from fixation)’ – the eccentricity of the flash stimulus with respect to the centrally-presented fixation cross; ‘1F0B (% correct or avg resp)’ – Accuracy scores (if presented as a percentage) or the mean number of flashes reported (if presented as a decimal number). Similar for remaining columns. Note for Innes-Brown et al (2011), ‘N’ and accuracy scores are given for both adults and children (adults, children).
3.2 Flash-Beep Illusion – First Reports (2000-2002)

The flash-beep illusion was first reported in 2000 by Shams and colleagues in a brief communication to the journal Nature (Shams et al., 2000). In this report the authors presented behavioural data indicating that when a single flash was accompanied by two or more beeps, the single flash was incorrectly reported as two or more flashes. The participants were presented with a number of brief flashes (varying from one to four on each trial) on a computer monitor. The flashes were either presented alone, or accompanied by one, two, three or four beeps. Participants were asked to count the number of flashes they observed.

When a single flash was presented along with two or more beeps, observers reported seeing two or more flashes. The number of flashes reported did not significantly differ from a control condition in which two real flashes were presented without concurrent auditory stimuli. The fact that subjects were able to accurately count three (and in some cases four) real flashes in the absence of any auditory stimulus indicated that the visual stimulus was relatively easy to perceive, and was not ambiguous. To assess whether participants were simply counting beeps rather than flashes, trials were presented in which a single beep was played along with the flashes. In these catch trials, subjects were also able to correctly report the number of flashes. The multiple beep stimuli thus caused a ‘fission’ of the visual stimulus. However, when multiple flashes were presented, the fact that subjects still reported the correct number of flashes when only one beep was presented indicated that the auditory stimulus did not equivalently ‘fuse’ visual stimuli. The authors proposed that their results are supportive of previous studies (Saldaña and Rosenblum, 1993, Gebhard and Mowbray, 1959, Shipley, 1964), suggesting the ‘discontinuity hypothesis’ – that the percept of a continuous stimulus in one modality is rendered more malleable by a discontinuous stimulus in another. Shams and colleagues also reported that the illusion remained strong when the auditory and visual stimuli were separated by up to 70-100 ms, a finding elaborated on in their next publication.
The next study published on the flash-beep illusion (Shams et al., 2002) contains much of the experimental data using the same protocols reported in the initial brief communication, and the data is followed by several hypotheses concerning the possible brain mechanisms behind the illusion. In one-flash trials, there were 0-4 accompanying beeps, and in multi-flash trials, there could be either a single beep or no beep.

The first beep always preceded the first flash by 23 ms. Each beep had a duration of 7 ms, and consecutive beeps were separated by 57 ms. The authors state that these timings were chosen in an attempt to match the beeps to the perceptual impression of the illusory double flashes, but that the timings were otherwise fairly arbitrary. In a personal communication with the author (7/04/2003), it was stated that “the timing of the visual stimulus was adjusted to give the real and illusory double flashes as similar count ratings as possible.” Another constraint on timing is the vertical refresh rate of the monitor used to present the stimuli. In this case the reported refresh rate was 60 Hz, giving rise to claimed flash duration of 17 ms (1/60). It should be noted that the real duration of the flashes is very likely less than the refresh period of the monitor, possibly of the order of 1-3 ms (Zele and Vingrys, 2005, Elze, 2010). As with many other reports of the flash-beep illusion, it is not stated whether the visual stimulus was displayed on a CRT or LCD screen. The implications of different visual display systems on stimulus timing are discussed further in Chapter 5.

The results in Shams et al. (2002) were in agreement with the initial report. The response data was presented in terms of the average number of flashes reported across all subjects. In single-flash trials, observers on average reported seeing just over one flash when a single beep was presented. However, when two or more beeps were presented along with the single flash, the average number of flashes reported increased significantly to a value just over two. Again, in the zero-beep trials, subjects were able to accurately count the flashes with the possible exception of the 4-flash trials, for which the mean response was approximately 3.6. These results indicate that by themselves, the flashes were relatively easy to count, for at least up to three flashes. In single-beep trials the pattern was very similar to
the zero-beep trails. The mean responses for each flash level closely followed the actual number of flashes presented. If participants were simply responding to the beeps rather than flashes, the mean responses would be close to one in the single-beep trials.

The authors then compared responses to the illusory double-flash trials to the real double-flash trials. The mean responses to 1F2B, 2F1B and 2F0B trials were on very similar points on a plot of the mean perceived flashes vs. stimulus type, but no statistical comparison was reported (see Figure 3-1).

![Figure 3-1](image)

Fig. 5. Comparison of different conditions in Experiment 1. The horizontal axis represents the number of beeps for the gray plot (corresponding to the single flash condition) and the number of flashes for the broken and solid plots (corresponding to the control conditions with a constant number of beeps). The overlap of all three plots at the second data point corresponding to one flash and two beeps for the gray plot and two flashes and one or no beeps for the solid and broken plots, respectively, suggests that the former condition is perceptually equivalent with the latter two so far as visual perception is concerned.

Figure 3-1. Average number of flashes perceived in Shams et al (2002). The red circle indicates the point where 1F2B, 2F1B and 2F0B data points are overlayed. The points are indistinguishable, indicating that responses to two real flashes alone, or two flashes with two beeps, are similar to when only a single flash is presented, along with two beeps. Figure and original caption adapted from Shams et al. (2002).
In a second experiment, the authors investigated the effect of separating the flash/beep pairs in time on perception of the illusion. They presented only the main fission illusion trial type – one flash paired with two beeps, and varied the time at which the pair of beeps started in relation to the single flash. The beep onset times used were 25, 70, 115, 160, 205 and 250 ms, both before and after the flash. They found that the strength of the illusion started to decline when the beep pair was offset by ±70 ms, but was still present at around ±115 ms offset. The effect was symmetric about the zero point – it didn’t matter whether the beep pair preceded or followed the single flash.

Taken together, the results from these first reports indicate that a stable visual percept can be altered by the presence of a concurrent auditory stimulus. While there have been other documented cases of audition influencing visual perception, these have mostly been cases where some qualitative aspect of the visual percept is altered. In the stream-bounce illusion (Sekuler et al., 1997) for instance, the trajectory of two balls (presented on a 2-dimensional display screen) on a collision course can be biased towards perception of a bouncing percept rather than a streaming percept by the addition of a sound played close to the moment when the balls intersect (Watanabe and Shimojo, 2001). In the flash-beep illusion, on the other hand, sound causes a categorical change in an otherwise easily-perceived visual stimulus. The illusion appears to be robust to training, and does not rely on special stimulus characteristics – either in terms of exact timing, brightness, or positioning (although it is stronger when the flash stimulus is in the visual periphery). When the timing of the stimulus display is altered so that auditory and visual stimuli are presented more than 75 ms apart, the illusion degrades, and disappears at about 160 ms.

### 3.3 Electrophysiological correlates

The results from the first two behavioural studies suggested that the illusion might reflect a widespread mechanism of multisensory integration in the brain, rather than accidental or marginal neuronal activity. The actual effect of auditory
stimuli on the visual processing pathway in the brain however remained unexamined. Shams et al thus investigated the effect of auditory stimuli on the flash visual-evoked potential (VEP), using a slightly altered version of the flash-beep paradigm detailed above. An observation in the discussion section of Shams et al (2002) noted that the illusion was stronger when the flash was presented in the peripheral visual field rather than centrally. This informal observation was tested (Shams et al., 2001) by presenting illusory (1F2B) as well as uni-modal stimuli (1F0B, 0F2B) both centrally and in the periphery, to determine whether the expected variation in performance could be correlated with neurophysiological measures. Event-related potentials (ERPs) were recorded, and compared in each location depending on whether the participant reported one or two flashes.

Subjects reported seeing two flashes on 81% of peripheral trials, 21% of the centrally-presented trials, and correctly counted two flashes in the real double-flash condition on 92% of trials. For centrally-presented trials, ERPs were calculated from trials in which a single flash was reported, and for peripheral trials, those where 2 flashes were reported were included for further analysis.

The ERPs were examined by calculating the $1F2B - (1F0B + 0F2B)$ difference wave for central and peripheral trials. This difference wave is commonly referred to in the MSI literature, and is usually written as $AV- (A + V)$. The amplitude of each point in the grand average waveform was then compared against zero. If more than 15 consecutive points (15 ms) were significantly different from zero, a multisensory interaction was deemed to have occurred.
Figure 3-2. Event-related potentials from Shams et al (2001). Significant multisensory interactions (1F2B – (1F0B + 0F2B)) are present at three occipital electrodes (Oz, O1, O2) from approximately 170-300 ms, only when the fission illusion stimulus is presented in the visual periphery. The top row shows responses when the stimuli were presented in the fovea, the middle row when the stimuli were presented in the visual periphery. The lowest row shows responses to a single flash subtracted from a double flash (ie leaving only responses to the second flash). Modified figure with original caption from Shams et al. (2001).

No interactions were found in centrally presented trials. In the periphery however, an ‘early’ interaction was found between 170 and 200 ms, and a ‘late’ interaction from 260 to 360 ms. The differences were consistent across the three occipital electrodes used in the study. The authors also examined the difference waveform calculated by subtracting the uni-modal single flash trials from double-flash trials, thus leaving a wave associated with the second flash alone. The results were very similar to the audio-visual interaction wave in the periphery, perhaps indicating that neural activity related to a second illusory flash is very similar to that related to the real second flash.
In summary, Shams et al (2001) reported a neurophysiological correlate of the illusory second flash which occurred at two distinct stages in the visual processing pathway, around 170-200 ms, and 260-360 ms. After Shams et al (2001) found that the VEP for an illusory flash is very similar to that obtained from a real flash, the data was re-analysed (Bhattacharya et al., 2002), this time investigating gamma-band activity in peripheral illusion trials (1F2B) trials depending on whether the subject indicated that they saw a single or double flash.

Figure 3-3. Time-frequency plots (average energy, inter-trial coherence and an interaction plot) for illusory vs non-illusory reports for 1F2B trials. Figure and original caption from Bhattacharya et al (2002).

Illusion and non-illusion trials were then compared with their respective baselines, or with the summed uni-modal responses, and subjected to a time-frequency transform to extract gamma activity. In illusion trials they found both an early (30-140 ms) and a late (420-480ms) increase in activity centred around 40 Hz. In non-illusion trials an early increase was found in the beta range, but no
corresponding late increase was found. The statistical significance of these differences (Wilcoxon signed-rank test) is indicated in Figure 3-4).

Figure 3-4. Statistical tests of gamma-band activity (compared to the pre-stimulus baseline) for illusory vs non-illusory reports for 1F2B trials. Figure and original caption from Bhattacharya et al (2002).

When inter-trial coherence (ITC) was examined in the same intervals, early gamma increases were found to be phase-locked to the stimulus for both illusion and non-illusion trials, but the later increase found in the illusion trials was not phase-locked to the stimulus (low ITC in the 150ms+ range). Standard averaging, as employed in (Shams et al., 2001), would thus not pick up this later increase.

A-V interactions were examined by calculating the AV-baseline – (A-baseline + V-baseline) TFR (Figure 3-3 panels e-f). The baseline in this case is the 100ms pre-stimulus period. Super-additive activity was found only in illusion trials, mainly before 200 ms, centred on around 40 Hz. When the 30-50 Hz band was isolated and analysed (compared with zero using Wilcoxon signed rank tests), regions of significant effects were found from 150-225ms and 425-500 ms).

In summary, Bhattacharya et al found that stronger early (30-140 ms) and late (420-480 ms) gamma-band responses, as well as supra-additive audio-visual interactions were correlated with perception of the illusion. The early gamma-band responses were time-locked to the stimulus, whereas the later ones were not.
3.4 Sound-induced fusion illusions

Andersen et al (2004) employed a modified version of the flash-beep task in order to more fully examine the stimulus parameters governing the illusion. They had an extended-factorial design, with 1-3 flashes and 0-3 beeps, as well as two sound intensity conditions – 80 dB, as in the original Sham et al studies, and a near-threshold condition where the intensity of the beep stimulus was dropped to 10 dB above threshold. They further divided the experiment into count-flash and count-beep blocks.

In their first experiment (count-flashes, 80 dB), the authors found the expected fission illusion: 1F2B was perceived as two flashes on 55% of trials, and in 1F3B trials, subjects reported three flashes on 24% of trials. These results are comparable to the initial reports from Shams et al (2002), although from experiment 2 in Shams et al (2002) the proportion of trials in which subjects perceive the illusion was higher (around 65%). Unlike the original experiments however, Andersen et al (2004) also found a corresponding fusion illusion. When stimuli with more flashes than beeps were presented, they found that participants tended to under-report the number of flashes they perceived. When two flashes were paired with one beep for instance, subjects reported seeing only a single flash on 57% of trials. Eight of their ten subjects showed significant fission illusions, and six out of ten showed fusion illusions. Overall the fusion effect was weaker than the fission illusion. When the participants were instructed to count beeps rather than flashes, it was found that the number of flashes had no effect on the number of beeps reported – the correct number of beeps was reported in nearly all cases.

In a second experiment, the intensity of the beep stimulus was reduced to 10 dB – a level that was only just perceived. While visual fission illusions persisted even with this very quiet sound, visual fusion illusions ceased to occur. In the count-beeps blocks, the 0-flash beeps were less accurately counted, and for trials in which a visual stimulus was present, the number of beeps reported was generally dependant on the number of flashes present. However the effect was strongest in
cases where the visual stimulus caused under-estimation of the number of beep stimuli. While all combinations of fission stimuli were significant, only the 1F3B fusion stimulus was significant.

In summary, although Andersen et al (2004) found fusion effects as well as fission, sound- and vision-induced fusion effects were weaker than fission effects, supporting the discontinuity hypothesis put forward by Shams et al (2002). The authors however state a weaker form of the hypothesis: that stimulus discontinuity is merely one factor among many that influences the tendency of one modality to dominate over another. In the near-threshold count beep blocks, Andersen et al found evidence for visually-induced illusory beeps – a finding that supports the information reliability hypothesis rather than the modality appropriateness hypothesis. Information reliability then is another factor influencing modality dominance. They also make the important observation that flashes and beeps must not always be integrated to a unitary percept, as the percept appears to have been very different in the count-beeps vs count-flashes conditions. Attention and task instructions appear to have a strong impact.

### 3.5 fMRI shows V1 activity is associated with the illusory flash

The electrophysiological studies of Shams et al (2001) and Bhattacharya et al (2002) provided neurophysiological support for the idea of ‘early’ integration between vision and audition by examining the timing of multisensory interaction effects on event-related potentials and fields. However, neither study could do more than speculate on the neural pathways that might be involved due in part to the limited spatial resolution of the techniques they employed. To address this, Watkins et al (2006) used fMRI to examine activity in visual cortex during perception of the flash-beep illusion. They used retinotopic mapping to isolate activity to the visual cortical area V1, and compared activity in this area during 1F2B trials depending on whether or not the illusion was reported.
Increased activity was found in V1, V2 and V3 during audio-visual stimulation compared to visual stimulation alone. When 1F2B trials were compared on the basis of whether or not an illusion was experienced, activity in V1 was found to be increased in illusion vs. no-illusion trials. This increase in V1 activity specific to perception of the illusion was further associated with enhanced activity in the superior colliculus and superior temporal gyrus.

Watkins et al (2006) related their findings back to a study that has shown that for near-threshold visual stimuli, increased activity in V1 is associated with successful detection (Ress and Heeger, 2003). They suggest that their own findings extend that result by showing that such an association of V1 activity with conscious experience extends to super-threshold visual stimuli under normal viewing conditions, and to changes in visual perception induced by an auditory stimulus.

Overall, the findings suggest that responses in V1 to a flash stimulus can be altered by sound, and furthermore that they are linked to subjective perception of the stimulus rather than the physical presence of the stimulus alone.

Watkins et al (2007) further suggested the possibility that the enhanced responses found in their previous study (Watkins et al., 2006) were due either to fluctuations in attention, or an automatic perceptual matching between the auditory and visual stimulus. To address this possibility, they repeated the study, looking at the ‘fusion’ illusion (Andersen et al., 2004). In 2F1B trials they found that 42% of participants reported seeing only a single flash. In these trials, brain activity in V1 was significantly lower than in the trials in which the real number of flashes was reported. As with their previous study, brain activity in 1F2B trials was again increased when two flashes were reported. Together, the two studies put forward a powerful case for the involvement of V1 in the perception of the illusory flashes - activity in V1 was increased during perception of two flashes when one was presented, and decreased during perception of one flash when two were presented.
3.6 The lasting effect of multi-sensory processes

The experiments employing the flash-beep paradigm reviewed so far have suggested that whatever mechanisms of multisensory integration underlie the illusion can operate at relatively early timescales (EEG studies), and that perception of the illusion can be correlated with increased activity in the early visual system (fMRI studies). This seems to be the case for both the classic ‘fission’ illusion, as well as the ‘fusion’ illusion. However the actual mechanisms behind how the illusion might operate remain unknown.

Meylen et al (2007) investigated the possibility that presentation of a multi-modal stimulus might affect subsequent visual processing (see Zangenehour & Zatorre (2010) for a recent example of this occurring over long time scales). The authors examined ERPs specific to the second flash, depending on whether the first stimulus in a pair was a uni- or multi-modal. In this way, they could avoid having to make use of the fission illusion stimulus – the 1F2B stimulus – at all. For instance, the 2F1B stimulus is re-cast as a “flash preceded by a multi-modal (flash/beep) stimulus”, and the 2F0B stimulus is “a flash preceded by a uni-modal (another flash) stimulus.”

The authors computed two difference waveforms designed to isolate activity common to the second flash. The UNI difference wave (UNI = 2F0B - 1F0B) isolated activity related to the second flash when followed by a uni-modal stimulus. The MULTI difference wave (MULTI = 2F1B - 1F1B) isolated activity related to the second flash when preceded by a multi-modal stimulus. Further contrasting UNI against MULTI then indicated that visual processing related to the second flash was differentially affected by the immediately preceding multi-modal compared to uni-modal stimuli.

Although the main interest of the study did not lie with the illusion trials, all six possible combinations of stimuli (1 or 2 flashes and 1, 2 or 3 beeps) were presented, enabling comparisons with previous flash-beep illusion studies. The results indicated that subjects performed extremely well counting the number of flashes in all stimulus combinations except the 1F2B ‘illusion’ trials – performance
very similar to that found in Shams initial report. Interestingly there is no sign of the “fusion” illusion reported by two other studies (Watkins et al., 2006, Andersen et al., 2004).

Figure 3-5. Behavioural results from Meylan & Murray (2007). Accuracy was very high in all trial types except the ‘illusory’ 1F2B trials. Figure from Meylan & Murray (2007).

In order to analyse the topography and sources of the difference, a multi-step procedure named ‘electrical neuroimaging’ (Michel et al., 2001, Murray et al., 2008) was performed. This procedure firstly analyses overall response strength and topography over time, in order to distinguish between ERP effects due to changes in generator strength versus changes in generator configuration. Secondly, the generators of these differences are estimated using source localisation techniques. ERPs were calculated for each condition, and the UNI and MULTI subtraction waveforms were computed. Changes in overall field strength were analysed by calculating the global field power (GFP - reviewed in Chapter 4). Figure 3-5 shows the UNI and MULTI ERPs from the study (Figure 3-5, panel A) as well as the MGFP. MGFP across all subjects differed between UNI and MULTI beginning 238 ms post-stimulus, or 173-210 ms after the onset of the second flash (Figure 3-5, panel B).

To statistically identify periods of topographic modulation (periods where scalp topographies are changing from one configuration to another), a measure of ‘global dissimilarity’ was calculated (Lehmann and Skrandies, 1980), and compared for UNI and MULTI waves. There was only one very short, and statistically insignificant period where the topographies of the UNI and MULTI stimuli
differed (Figure 3-6, panel C). An analysis of the topographic maps then revealed a stable map topography in the time period 225-297 ms, overlapping with the difference in GFP between UNI and MULTI ERPs.

Figure 3-6. Top/Left: Comparison of the UNI (labeled “VIS” above) and MULTI waves revealed an attenuation of the MULTI wave compared with the UNI wave for a short interval from 238-275ms after the first flash. Bottom/Right: Estimated sources of the UNI and MULTI difference waves (and the mean difference between them) in the 225-297 ms time period. Modified figure adapted from Meylen and Murray (2007).

In the final step of the analysis, the source generator of the ERP in this period was estimated using the LAURA technique (Grave de Peralta Menendez et al., 2001, Grave de Peralta Menendez et al., 2004). Similarly to the ERP data,
activity was reduced in the MULTI condition compared to VIS, but there were no
differences in location of the sources between the two conditions. Activity in both
conditions was mainly focussed in the occipital and inferior parietal lobes on both
sides of the head. The difference map indicated that the largest difference between
the two conditions was localised to the left occipital lobe (Figure 3-6, lower right
panel).

The results thus indicated reduced activity at about 160ms (relative to the
second flash) in the occipital lobes to a visual stimulus when it was preceded by a
multisensory (audio-visual) rather than a uni-sensory (visual) stimulus. Although
the source localisation procedure indicated areas of the brain that are usually
considered primary sensory areas, the timing of this difference is not consistent
with low-level responses. The authors thus suggested their data was generally
supportive of the hypothesis that low-level stages of the visual cortex are involved
in multisensory processes, but do not postulate on whether this could be due to
direct input from the auditory system to V1, or due to feedback within the visual
system.

The only significant effects found in this study were in terms of the GFP.
No differences in the topographic patterns were found between VIS and MULTI
stimuli. This is surprising - if anything one might predict different activity in the
MULTI condition, if auditory inputs were being directed to primary visual areas.
The most straightforward interpretation of GFP differences in the presence of no
map differences is in terms of a difference in the strength of activation of a
network of indistinguishable sources – or that the attenuation in the MULTI
condition occurred in brain regions that were already active in the UNI condition.
The fact that there were no differences in the sources observed also supports this
idea.

To put these results in the context of other studies, Watkins et al (2006)
looked at activity in V1 in response to 1F2B stimuli, depending on whether or not
the illusion was perceived. They found that activity was increased if the extra flash
was perceived. When an auditory stimulus was preceded by a multi-sensory
stimulus, there was an increase in V1 activity (for trials when the illusion was perceived). Correspondingly in Meylan at al (2007), activity was reduced when a visual stimulus was preceded by a multisensory stimulus. In this case however, there was no corresponding alteration in perception. So multisensory stimuli affect subsequent stimuli differentially depending on their modality – a following beep is associated with increased activity in V1 (and a change in perception), whereas a following flash is associated with a reduction in activity in V1, and no change in accuracy of perception.

### 3.7 Further psychophysics (2007)

Against the background of gradually more complex neural investigations of the most basic form of the illusion, only two studies have more closely examined the basic psychophysics of the illusion. The first study (Courtney et al., 2007) followed up the original contention that the illusion can be explained by the discontinuity hypothesis (Shams et al., 2002). In their original report, Shams et al (2002) found that although a double-beep could cause the perception of a double-flash, the reverse was not the case – a double flash did not cause the perception of a double-beep. They thus dismissed the modality appropriateness hypothesis in favour of an explanation based on their observation that it was the transient or discontinuous stimulus that dominated the overall percept. Courtney et al (2007) tested whether this discontinuity hypothesis could be extended to concurrently-presented stimuli (the auditory and visual stimulus in Shams et al were not presented with simultaneous onsets), and whether it operated within as well as across modalities. They found that the double-beep stimulus still caused a single flash stimulus to be perceived as two when the stimuli were presented concurrently, and to a slightly lesser extent when a much longer (317 ms) flash stimulus was used. Interestingly, their mean response (number of flashes perceived) for the 1F2B stimulus was almost 2.5 – *more* than the number of beeps presented. Although multi-flash trials were included in a second experiment, the longer 317 ms flashes were divided into two by the insertion of a single blank frame, leaving two 150 ms flash stimuli with a 17 ms gap, resulting in a paradigm
significantly different to that used in most other studies reporting the illusion. When they included this long double-flash stimulus, a single beep did not appear to affect perception of the double flash, again supporting the discontinuity hypothesis. However, the use of such a long flash stimulus makes the results difficult to compare to other studies. A third experiment tested whether an illusory flash could be elicited by another discontinuous visual stimulus. Short (17 ms) single and double flash stimuli were presented near the fovea (2° above fixation) and peripherally (5° below fixation), with subjects reporting how many flashes they saw at the lower peripheral position. They found that when a single flash was presented at the lower location, subjects reports increased from one flash to two flashes as the number of flashes at the higher foveal location increased from one to two. When two flashes were presented at the lower peripheral location, subjects reported a double flash when either one or two flashes were presented at the top location. In other words, they found that the double-flash illusion could indeed be elicited in the periphery by a foveal visual stimulus. Again, they did not find that a single foveal flash would fuse a double peripheral flash, with subjects accurately reporting two flashes in that case.

In summary, Courtney et al (2007) found that the flash-beep fission illusion persists with concurrent auditory and visual stimuli, and with relatively long visual stimuli. They also found that the flash fission in the periphery could be elicited by another flash presented at the fovea. They cautiously suggest these results are all supportive of a general discontinuity hypothesis that may not be bound by stimulus modality or structure. However, their lack of finding a corresponding “fusion” illusion (as reported by Andersen et al, 2004) cannot easily be compared to other studies due to the very long visual stimuli used.

Evidence from EEG (Bhattacharya et al., 2002, Shams et al., 2001), MEG (Shams et al., 2005a) and fMRI (Watkins et al., 2006, Watkins et al., 2007, Zhang and Chen, 2006) studies have all shown neurophysiological correlates of the illusion in primary visual areas, thus giving weight to the original suggestion that the illusion is perceptual rather than a result of response biases. However only one study has directly attempted to establish whether the sound merely creates
confusion, rather than evoking an illusory flash with visual characteristics similar to a genuine visual stimulus.

McCormick and Mamassian (2008) presented the flash-beep paradigm concurrently in two locations (left and right of fixation). In their first experiment, high contrast flashes (uniform white box against a grey checker-board background) were presented to both locations, followed by a second low-contrast flash (uniform grey box against the same background) to both locations on half the trials. Either a single or double beep was also presented on each trial. Observers reported whether a low-contrast flash was perceived on each trial. Sensitivity (d’) and criterion (c) measures were calculated, and showed that in double-beep trials, participants showed a high false alarm rate (associated with illusory perception of a second flash) as well as reduced visual sensitivity.

In a second experiment the authors aimed to determine whether the illusory second flash had a detectable contrast. In this experiment the initial high-contrast flash could either be black or white, and was followed by a dark or light low contrast flash on only one side. A staircase procedure was run to determine the second-flash contrast required to obtain 75% accuracy on a location detection task where participants were required to report the side of the second flash. Half the trials had single, and half double beeps. Comparison of the contrast thresholds revealed that with a white high-contrast flash, the presence of two beeps facilitated the detection of a real light low-contrast flash, but this did not hold for the darker low-contrast flashes. When the initial high-contrast flash was black, the double beep did not alter contrast detection thresholds for either the light or the dark low-contrast flash.

By showing that contrast detection thresholds were lowered for light but not dark second flashes, the authors infer that the illusory flash, which supposedly occurs at the same time as the real second flash, has interfered with perception of the real second flash. In the case of the light flash, it lowered the contrast required to perceive the second flash, making it easier to perceive. With dark second flashes, the contrast detection threshold did not change. The inference then, is that
with a white initial flash, the illusory flash is of the same contrast polarity. The authors suggest their data is supportive of the idea that the illusion is a perceptual phenomenon, as opposed to merely a reflection of response bias induced by sound, and that the illusory flash has psychophysically measurable characteristics. It is unclear why perception of the dark second flashes was not facilitated by the black initial flashes. Perhaps an illusory flash is always of light contrast, no matter what the initial stimulus.

### 3.8 Advanced Electrophysiology (2007-2009)

Mishra and colleagues have published a series of three electrophysiological investigations of the flash-beep illusion (Mishra et al., 2007, Mishra et al., 2008, Mishra et al., 2009). In all of these studies, an extended selection of ten stimuli were presented. In addition to the usual six stimuli (1 or 2 flashes, with 1, 2, or 3 beeps), Mishra and colleagues presented auditory-only stimuli (0F1B, 0F2B), a 1F2B condition with a delayed flash (200 ms after the beeps) and a no-stimulus condition. Activity related to the illusory flash was then computed by subtracting the waveform for the added uni-modal stimuli from the multi-modal stimulus (interaction = AV – (A + V).

In the first study (Mishra et al., 2007), the amplitude of an early positive component in the AV difference wave, peaking at 120 ms (PD120), was found to vary depending on the frequency with which individual subjects tended to report the presence of the illusion, and was completely absent in participants who never reported the illusion. The component was localised (using the same LAURA technique used in Meylan & Murray (2007)) to the extra-striate visual cortex. Within subjects, however, the PD120 amplitude was not predictive of whether an individual reported the illusion or not on any given trial.

In the second study (Mishra et al., 2008), similar results were found for the equivalent fusion illusion. In this case, a component in the fusion AV-difference waveform (PD180) was much smaller in participants who more frequently reported the fusion illusion. Unlike the case with the fission AV difference wave,
the PD180 component was located to superior temporal areas, and the amplitude was found to correlate with participants reports of the illusion on individual trials – on trials when the illusion was present, the PD180 component was significantly smaller. Taken together, the results from the experiments suggest that the fission and fusion illusions are each related to brain activity with quite different timing and spatial location, and that top-down cognitive factors such as attention can influence the illusion.

In the third study (Mishra et al., 2009), the effect of spatial selective attention on these processes was examined. Despite the fact that most of the previous electrophysiological studies of the flash-beep illusion have tended to suggest involvement of early, pre-attentive mechanisms in the brain, few studies have specifically tested the effect of attention on the illusion. Selective attention is known to improve performance, both in terms of detection and reaction times, to classes of stimuli to which it is allocated (Luck and Hillyard, 1994b, Posner et al., 1980). The behavioural advantage of attention has been related to increases in neural activity related to processing attended vs unattended stimuli at a variety of levels. ERPs are generally enhanced to attended stimuli, with effects starting after about 100 ms post-stimulus. In the case of the flash-beep illusion, this general pattern appears to hold. When the stimuli were presented simultaneously in both the upper and lower visual fields, all early ERP components in the crossmodal difference wave associated with the illusion were enhanced by selective spatial attention.

### 3.9 Applications – Development (2007)

Finally, the flash-beep illusion has been used as an index of multisensory integration processes in childhood development (Tremblay et al., 2007), and more recently in autism research (Foss-Feig et al., 2010) and elderly populations (Setti et al., 2011). Only the developmental literature will be reviewed here, although the principles of using the illusion in any population are the same – the presence of the illusion is assumed to indicate an increased susceptibility to the illusion. Reports
of the illusion indicate the presence of multisensory processes, but they also indicate that the process itself is immature, or in some way impaired compared to a control population. The development of audio-visual integration in humans was reviewed in Chapter 3, however, developmental studies with specific focus on auditory-visual illusions will be briefly noted here.

Studies examining the McGurk illusion in children and adults have reported fewer perceived illusions (instances of multi-sensory integration) in younger participants and found that when responses were dominated by a single sense modality, children relied more heavily on the auditory input (McGurk and MacDonald, 1976, Massaro, 1984). Although several studies have shown the McGurk illusion to occur in infants (Rosenblum et al., 1997, Burnham and Dodd, 2004) and children (McGurk and MacDonald, 1976, Massaro et al., 1986), there is currently only one study that has examined the developmental time-course of multisensory integration using simple, non-speech stimuli (Tremblay et al., 2007).

Tremblay et al (2007) examined performance on the McGurk task, as well as both fusion and fission illusions arising from the flash-beep task. Results were compared across three age categories (5-9, 10-14 and 15-19 years old), thus giving speech-related and non speech-related metrics of MSI across three age groups. They found that there was a relationship between age and performance on the McGurk task, with 5-9 year old children reporting significantly fewer McGurk illusions than both of the older groups, and a significant correlation between the number of trials with McGurk illusions and age. No age relationships were found on either the fusion or fission flash-beep illusions however. One interpretation of these results might suggest that the developmental time-course of MSI is different, depending on whether the task involves speech, or simple non-speech stimuli. Along with the previously-described studies of the McGurk illusion, these results generally suggest that the influence of visual stimuli on multi-sensory processing increases with age for speech stimuli, but not for non-speech stimuli. However, as the authors note, it is important to recognize that these tasks differ in more than just the presence or absence of speech stimuli. For instance, in the McGurk task participants are asked to make an auditory judgment, while in the flash-beep task
the task is a visual judgment. It is also interesting to note that the accuracy scores in fission illusion trials for all three groups of children are noticeably lower than those found in studies of the fission illusion in adults (see Table 3-1). Although there is no developmental trajectory evident across the three age ranges tested in the study, it is still possible that the children were seeing more fission illusions than adults. It is difficult to assess this possibility further without testing a group of adults and children using the same paradigm however (the experiment described in Chapter 7 overcomes this problem).

3.10 General Discussion

The flash-beep illusion has been used as an index of multisensory processes in a number of areas. Psychophysics experiments have suggested that the illusory second flash is indeed perceived as a ‘flash,’ with describable visual features such as contrast. Others have begun to explore the parameter space of stimuli that will support the illusion. For instance, it has been reported that the illusion is stable with a variety of audio-visual stimulus timings, and different shapes and types of visual stimuli. However, there have been no reports as yet describing how separation of the stimuli in space affects perception of the illusion. EEG and fMRI studies have also begun to discover the neural foundations of the illusion – for example areas of primary visual cortex become more active when the fission illusion is reported, and less active when the fusion illusion is reported. The timing of the illusory effects as well as the results from neurophysiological studies have suggested the involvement of direct, cortico-cortical connections between primary auditory and visual sensory cortex. The functional properties of these connections fit well with behavioural features of the illusion: As the illusion is known to occur more strongly in the visual periphery, and the connections recently found between A1 and V1 in primates are known to project more strongly to the peripheral visual field, separating the auditory and visual components of the stimuli may provide further insight into the involvement of these connections. The experiment reported in the next chapter separates the flash and beep stimuli by 20 degrees in the visual field, divided across the midline.
4 EEG Analysis Techniques
4.1 Discovery of scalp potentials

In 1875, Richard Caton, using a light-reflecting galvanometer with two wire electrodes, reported to the 43rd annual meeting of the British Medical Association that electrical currents could be recorded from the exposed brain surface of rabbits, cats and monkeys (Caton, 1875). Caton found that the beam reflected by the galvanometer moved while the electrodes were either both placed on the gray matter surface, or with one electrode on the gray matter surface and the other on the scalp. He also made the important observation that the surface of the gray matter was positive with reference to the deeper structures, and coined the term ‘negative variation’ to indicate when this positivity was interrupted by functional activity. However it was not until 1924 at the University of Jena that neuropsychiatrist Hans Berger first detected very small currents from the human scalp – using at first a series of string galvanometers, and later a more sensitive double-coil galvanometer. Used with non-polarisable pad electrodes, this apparatus had a sensitivity of around 130 μV/cm (Niedermeyer and Lopes da Silva, 2005). From his experiments he observed that regular oscillations in the brain were attenuated during sensory stimulation, and divided the oscillations by their frequency into alpha (~10 Hz) and beta (~4 Hz) bands. At first Berger was sceptical that the oscillations he observed had cerebral origins, and so set about a series of experiments correlating his measurements with mental functions such as sleep, coma, and attention. He finally published these results in 1929 (Berger 1929). Interestingly, he hypothesised that alpha rhythms represented a form of automatic processing that was interrupted (or desynchronised) by attention or external stimuli. The modern concept of event-related synchronisation and desynchronisation is very close to this idea and will be discussed later in this chapter. Throughout his life Berger was fascinated by the mind/body problem, and worked towards the rejection of dualism as an explanation for consciousness. He saw the measurement of brain electrical activity as a prospective means of clarifying this relationship, and set about finding a physiological measure of brain activity that was related to mental processes – a task that many cognitive scientists are still
pursuing to this day. In this respect his discovery of the human EEG *per se* was not the main goal of his activities, however each of his experiments examining how the EEG changes during sleep, hypoxia, epileptic seizure, and external stimulation were all ground-breaking, and provided much of the basis for modern cognitive electrophysiology.

Despite the depth and breadth of Berger’s experiments, his findings were largely ignored by the scientific community until they were replicated by Lord Adrian and Brian Matthews at Cambridge (Adrian and Matthews, 1934). Adrian used multiple electrodes to show that Berger’s alpha rhythm arose from occipital regions of the scalp, and that a train of “evoked potentials” could be recorded if a visual stimulus in the whole visual field was made to flicker. Furthermore, changes in the rate of the flicker stimulus were found to be reflected in the rate of the occipital potentials (Fig 2-1). Although not averaged event-related potentials in the sense we know them today, these trains of brain potentials were the first sensory-specific evoked potentials to be recorded in humans. Figure 2-1 shows EEG traces during flicker stimulation from Adrian & Matthews (1934) alongside the output from a photodiode measuring the flicker rate – as the flicker rate is reduced from 18 to 8 Hz, the frequency of the EEG can clearly be seen to follow the flicker rate.
Figure 4-1. An example of Berger Rhythms following the rate of a flicker stimulus. Figure and original caption from Adrian & Matthews (1934), *Brain*, 57(4), pp 379.

The first auditory potentials were described in 1939 (Davis, 1939). Diphasic or triphasic waves, with negative initial phase, were recorded in response to loud and faint tones from 200 to 2000 Hz. The response was largest at the vertex (compared to occipital, frontal and temporal regions), and had a latency of around 100 to 150 ms (measured from Fig 1, panel C in Davis, 1939, reproduced as Figure 4-2 in this manuscript). Figure 4-2 shows the onset and offset responses to 250 Hz and 2000 Hz tones in a number of experimental participants. This vertex potential was then shown to be modality non-specific, appearing in response to tactile, visual, and electric shock stimulation to the finger, and was much larger during sleep (Davis et al., 1939). The small size and large trial-to-trial variability of the vertex potential made repeatable investigations difficult however, and very little further research initially followed from these results.
It was not until 1953 that the British medical scientist George Dawson gave the field its next big advance. During a study of cortical waves related to myoclonic seizure, he noticed that these relatively large waves were more easily examined if several sweeps were photographically overlaid, rather than examining each sweep separately (Dawson, 1947b). In normal subjects however, these waves were less than a quarter of the size of those from the patients with schizophrenia (Dawson, 1947a) and much more difficult to discern with superimposition – clearly a more sensitive approach was called for. Dawson’s answer to this problem was the automated signal-averaging machine. His first machine was an electro-mechanical device with two recording channels, and was a fascinating piece of engineering though a full description is sadly beyond the scope of this thesis.

Using this machine, Dawson could easily detect small voltages that would normally

---

be buried in the on-going EEG. Figure 4-3 shows averages from 55 trials made with the electro-mechanical averager. On the left are the individual trials from an electrical stimulus to the left wrist, and on the right is a 5 µV calibration pulse, again with 55 trials. The top panels show the superimposed trials, and the bottom panels the averaged response. Echoes of this superimposition technique are now re-appearing in the form of the ‘ERP image’ (Makeig et al., 2004). Although the electro-mechanical averager was a huge step forward, it was not until an electronic version (Dawson and Pitman, 1958), that the way forward to mass-produced computerised averaging was realised.

Figure 4-3. Averages from 55 trials made with the electro-mechanical averager. On the left are the individual trials from an electrical stimulus to the left wrist, and on the right is a 5 µV calibration pulse, again with 55 trials. The top panels show superimposed trials, and the bottom panels the averaged response. Figure and original caption from Merton & Morton (1984).
Electronic averaging over a large number of trials proved to be an extremely successful technique, allowing precise timing of brain responses in relation to external stimuli, and even in relation to behavioural responses. At first the averaged responses were termed “evoked potentials”, however as it was gradually realised that brain potentials could be generated in response to internal or cognitive events, the more inclusive term “event-related potential” (ERP) began to be used. Measurement of the averaged ERP allowed researchers to develop an almost infinite variety of experiments testing brain responses to every imaginable stimulus type, and allowed the first non-invasive “window” into the workings of the intact, performing human brain³. Over the years, amplifiers have dramatically shrunk in size and increased in the resolution of analogue to digital conversion, the number of recording channels, and sampling rates available.

The development of off-the-shelf EEG recording and analysis packages in the early 1960’s allowed more laboratories to begin studying ERP’s, and by the end of the 1960’s the previously-intractable “vertex potential” was disambiguated into a series of negative and positive deflections that differentially responded to experimental manipulations (Davis and Zerlin, 1966). Presently two systems of nomenclature are in use – the deflections are either coded by polarity and sequence (P1, N1, P2 and N2) or by polarity and latency (P50, N100, P175 for example). As ERP peaks can have considerable variation in latency with different task or stimulus conditions, the former approach (polarity and sequence, ie N1, P2 etc) will be used in the present work, unless clarity in quoting the work of others necessitates otherwise.

Before reviewing the basics of auditory, visual, and multisensory ERP’s, it is important to point out a distinction that is often lost – that of the difference

³ Since 1960, the number of reports using this methodology has exploded almost geometrically, with 4299, 13753, 24555, 29227 and 34065 new reports published in each of the decades ending 1970, 1980, 1990, 2000 and 2010 respectively. The search term submitted to PubMed for this search was “evoked potential” OR "event-related potential" OR ERP”, and was performed in the middle of 2010 (the “decade” ending 2010 was as yet incomplete).
between the maximum or minimum coordinates of a deflection on the EEG, and
the concept of a “component”. Although the peaks and troughs of the ERP are
convenient points for measurement, each individual peak is not necessarily related
to an individual event or generator in the brain. The latency and morphology of
each peak is multiply determined – by the location of the scalp and reference
electrodes, and by the latency and location of possibly many brain processes
summing to produce a single signal at a particular electrode. It is tempting
however, to refer to a particular peak (the N1 for example) in the averaged ERP,
as the N1 “component”, thus inferring the peak itself is a direct index of some
proposed generating process in the brain. In fact, a component should more
properly be described as “the contribution to the recorded wave-form of a
particular generator process” (Naatanen and Picton, 1987). Many components
contributing to the auditory N1 have now been identified, so that the N1 is itself
actually composed of multiple overlapping components, representing activity in a
variety of cortical sources at a variety of times. For the N1, the current consensus
is at five subcomponents per hemisphere (Hari et al., 1982, Picton et al., 1999).

4.2 Generation of the EEG

The measurement of a potential difference between two points on the scalp
is merely a gross indication of current flow somewhere within the skull. The brain
is made up of around $10^{11}$ nerve cells, around $10^{12}$ glial cells, and associated blood
vessels and membranes (Kandel et al., 1991). Figure 4-4 depicts a typical cortical
pyramidal neuron. From the main cell body emerge the tree-like apical dendrites,
which receive from $10^4$ to $10^5$ synaptic inputs from pre-synaptic neurons, and the
axon, which carries the signal to other neurons. There are two main types of
electrical activity associated with neurons – action potentials and post-synaptic
potentials. Action potentials have very short durations (microseconds) and hence
neighbouring neurons are likely to produce destructively-cancelling fields. The
generation of current flows detectable at the scalp is thus thought to be a
consequence of both inhibitory and excitatory post-synaptic potentials (IPSPs and
EPSPs) in cortical pyramidal neurons (Nunez and Srinivasan, 2005).
In the case of an IPSP at the cell body, current flows across the synaptic membrane, through the extracellular fluid, across the dendritic and axonic membrane, and back from the basal dendrites to the cell body through the intracellular fluid. For an EPSP, this current flow is reversed. The flow of current away from the apical dendrites into the cell leaves a net negativity on the outside of the apical end of the cell. Likewise, the return flow results in a positivity at the basal end of the cell. This graduated charge along the length of the cell constitutes a tiny dipole source. When the dipoles from many neighbouring cells overlap and align, the resulting net dipole (called the equivalent current dipole or ECD) is measurable as a voltage difference between two locations on the scalp.

Figure 4-4. The generation of scalp EEG. From Luck and Girelli (1998). Note that the location of the dipole (immediately below the vertex) is very different to the scalp distribution (skewed to the right), because the dipole is located within a sulcual fold, and thus not oriented tangentially to the scalp. Because of this, scalp maps are not necessarily a good indicator of neural sources.

Spatially coherent arrangements of cells are critical for the superposition of neural currents such that they overlap to produce measureable fields at the scalp.
(Figure 4-4, panel B). This arrangement is present in the cortical pyramidal cells, which are arranged perpendicularly to the cortical surface. The currents associated with EPSPs along an individual dendrite are believed to form a current source around 20 fA-m in strength, far too small to measure at the scalp. Rather, empirical observations suggest sources of around 10 nA-m (Bailet et al., 2001). Taking into account the density of neurons in the cortex this yields a cortical current density of around 100 nA/mm$^2$ (Hämäläinen et al., 1993). Assuming a cortical thickness of 4mm, a patch of cortex measuring around 5 mm x 5 mm would thus yield a net current of 10 nA-m, consistent with measurements (Hämäläinen et al., 1993).

The conduction of the ECD to the scalp surface occurs by a process called volume conduction, illustrated in Figure 4-4, panel C. Volume conduction describes the way that electric fields are distorted by the different conductivities of the various tissues of the head. The gray matter, white matter, dura, bone and skin all have very different electrical properties. As a result the fields measured at the scalp surface are very much blurred, or spread out across the head. The spatial resolution of the EEG is thus fairly limited.

### 4.3 Auditory ERPs

Unlike the visual ERPs which are generated mainly in the cortex, it is possible to record auditory ERPs that arise from various locations in the brainstem. These auditory brainstem responses (ABRs) occur in the first 10 ms after a stimulus. These very early responses will not be reviewed here, but see Nunez & Srinivasan (2005) for a review. In adults the auditory ERP consists of a series of positive and negative deflections, with the first positive peak at around 50 ms (P50 or P1), a large negative peak at around 100 ms (N1 or N100), and a second positive peak at around 200 ms (N2 or N200).
Figure 4-5. A cartoon representation of the various component waveforms of the auditory event-related potential. Waves I-VI are generated in the brainstem. Waves P1/P50, N1 and N2 are generated in the cortex but are usually considered ‘obligatory,’ or mainly stimulus-driven, and N2 onwards are ‘endogenous,’ or more related to cognitive events. Figure from (Rissling and Light, 2010).

These three main deflections occur irrespective of most task demands, although experimental manipulations do have effects on their size and latency. Depending on task demands there may be other later components that follow or overlap (mismatch negativity, P300), however the experiments in this thesis are mainly concerned with the early stages of stimulus processing, so these later components will not be reviewed here. The amplitude and morphology of the N1 is highly dependant on the acoustic parameters of the stimulus used, the interval between stimulus presentations (Budd et al., 1998), and attention (Hansen and Hillyard, 1983, Alho et al., 1987, Naatanen and Picton, 1987, Naatanen et al., 1993, Teder et al., 1993).

The N1 is generated by at least three bilateral sources in the auditory cortex. The first of these is in PAC (see Chapter 2), located on the supratemporal
plane of the temporal lobe, or Heschl’s gyrus (Pantev et al., 1988, Pantev et al., 1991, Zouridakis et al., 1998). The tangential orientation of this source results in a fronto-central scalp distribution. The second component is observed at mid-temporal sites, with a radially-oriented dipole source also located in the planum temporale (Lutkenhoner and Steinstrater, 1998, Godey et al., 2001). There are also thought to be generators in frontal areas (Naatanen and Picton, 1987, Naatanen et al., 2011).

4.4 Visual ERPs

The visual ERP (sometimes referred to as the visual evoked potential or VEP) usually takes the form of at least three deflections – C1, P1, N1, N2. P2 and P3 may also be apparent depending on task and stimulus conditions. C1 is largest at posterior midline sites, occurs at around 40-100 ms, and its polarity varies depending on the location of the stimulus in the visual field (Di Russo et al., 2002). The C1 is sensitive to basic stimulus parameters such as contrast and spatial frequency (Clarke, 1973), and is thought to be generated in V1, an area of cortex that is folded into the calcarine fissure. The lower half of the visual field is represented on the upper half of the fissure, and the upper visual field in the lower half of the fissure. As a result, a stimulus presented in the upper visual field will result in a negative C1 at scalp electrodes above the fissure, and vice-versa for stimuli presented in the lower visual field (Di Russo et al., 2002).

The next peak, the P1, occurs around 90-130 ms. The P1 is sensitive to stimulus factors such as contrast and spatial frequency, but can also be affected by the direction of attention in space (Hillyard et al., 1998). Depending on the presence and polarity of the preceding C1, the P1 onset time and peak time can sometimes be difficult to assess. Following the P1, the N2 peaks at around 100-150 ms anteriorly, and a little later at posterior electrodes. The N2 is also sensitive to spatial attention (Hillyard et al., 1998). The visual N2, similarly to the auditory N1, is composed of several subcomponents. The lateral occipital component is relatively larger during discrimination compared to detection tasks (Vogel and
The P2 wave then peaks following the N1. The P2 is larger to target stimuli in oddball tasks, but only when the target is defined by simple features (Luck and Hillyard, 1994a). The P2 is often overlapped by the N1, and P3 waves if present, and is probably the least-studied of the visual ERP features.

### 4.5 Multisensory (auditory-visual) ERPs

When a multi-modal stimulus is presented, the scalp ERPs can become difficult to interpret. In the case of an audio-visual stimulus, both auditory and visual ERPs will be elicited, and will constructively and destructively interfere at different time-points and at different locations on the scalp, in a manner dependant on the configuration of the measurement and reference electrodes. In many ERP experiments on multi-sensory interactions, auditory (A), visual (V), and auditory-visual (AV) stimuli are presented during speeded detection or discrimination tasks. In these experiments, non-linear response interactions have been reported at very early time periods (around 50 ms) during detection and discrimination tasks (Giard and Peronnet, 1999, Molholm et al., 2002). Findings such as these do not rely on the measurement of pre-defined ERP components, rather the approach is to search for differences between conditions across the entire waveform, controlling for the large number of multiple comparisons, and then linking these differences to structures in the brain by inference from either the timing and electrode locations, or by localisation of the electrical sources of the differences. This approach is particularly suited to designs incorporating multi-modal stimuli, as no prior assumptions about time periods or pre-determined ERP components are necessary. A fuller description of how this technique will be applied in this thesis will follow in Section 4.7.

### 4.6 Artifacts and Independent Components Analysis

As well as recording electrical activity from the brain, electrodes on the scalp pick up activity from various other sources. Some of these sources are also
biological, and come from muscle activity or eye-movements etc, and others originate from external sources, such as oscillating electric fields caused by fluorescent lights or power supplies nearby to the recording equipment. Various steps can be taken to control the external sources, such as ensuring that electrode impedances are low and equal across the scalp, electrode lead wires are kept away from power cables, and that no large transformers or motors are too close to the recording equipment. Other biological sources, however, can be more difficult to control, as they cannot be simply removed from the room or shielded. Biological sources include things such as DC offsets caused by sweat on the skin (which can be controlled to some extent by maintaining a cool and comfortable room temperature) and muscle activity, which can either be related to continuous tension, especially in the muscles of the neck and face, or movements (talking, coughing, general fidgeting). One of the biggest problems, however, are the large, low-frequency signal excursions caused by eye-blinks.

There are many existing techniques for dealing with artifacts in EEG data, ranging from simple voltage-threshold based rejection to sophisticated and automated methods (for reviews see Croft and Barry, 1998, Croft et al., 2005, Schlögl et al., 2007). In this thesis, however independent components analysis (ICA) was used. This technique has the advantage that it does not require ‘templates’ of artifacts to be corrected, and can correct a wide variety of artefactual activity. The mathematics of ICA analysis are complex and will not be explained in full here (see Jung et al., 2001, Makeig, 1996, Onton et al., 2006 for full details of the mathematics involved). Briefly, however, ICA is an example of a blind source separation algorithm, a class of algorithms originally proposed to recover independent sources from noisy signals, such as recovering a voice from a noisy audio recording. More recently, it has also been used in a similar way to remove noisy sources from multi-channel EEG signals. ICA finds an un-mixing matrix, which linearly decomposes the multi-channel EEG into the sum of a number of maximally temporally independent and spatially fixed components. The number of these components is equal to the number of EEG channels.
Each component is associated with a scalp topography (spatial features) and time course (temporal features), which represents the relative projection strength of that component across every scalp electrode. By examining the temporal and spatial features of each component, it is possible to determine which components are likely related to brain sources (they tend to have smooth, dipolar maps, and timecourses with spectrums in the normal physiologically-plausible ranges), and those which may be related to external sources (such as a timecourse with spectral peak of 50 Hz across all electrodes – likely to be related to electrical interference), or components with maps focussed on the frontal poles, with very low frequency timecourse spectral peaks (likely to be related to eye-blinks).
Figure 4-6. Time-courses of all independent components calculated for subject N019 in Chapter 6. Components 1, 4, 10, 14, 29, and 40 were marked for rejection in this case. Component 1 appears to be related to blinks, and the remaining marked components are related to occasional electrode problems or high frequency (50 Hz) noise.
Figure 4-7. The scalp map, ERP image, and power spectrum of component 1 for subject N019. This component was marked for rejection as it appears to be related to eye-blink activity.

Figures 4-6 and 4-7 illustrate the use of ICA to identify an eye-blink component. Figure 4-6 shows the timecourses of all the components calculated from a representative subject from the experiment in Chapter 6. The first component has a timecourse featuring occasional low-frequency, large-amplitude excursions. Figure 4-7 (upper left panel) shows the scalp map, indicating that this component is most heavily weighted to the electrodes at the frontal poles. The spectrum (lower panel) shows that the timecourse activity has a large spectral peak at around 2-3 Hz. Lastly, the upper left panel shows how the timecourse activity is related to the epoch timing. The surface plot shows every trial ‘stacked,’ from the first trial at the bottom of the plot, to the 100th trial at the top of the plot. The colour intensity shows the voltage. In this case, the epoch was stimulus-locked, and as is sometimes the case, it appears that the participant blinked approximately 500 ms after nearly every stimulus. Using a voltage rejection criteria to address this problem would clearly result in the loss of nearly all the experimental trials for this participant.

Once the artifactual components are identified, they are marked, and all time points are set to zero. Using the electrode location data, the components are
then projected back into the electrode space, and the user is left with EEG data in
the usual format (voltages as a function of time for each electrode), but without
the contribution from the artifactual components. As will be appreciated, there is
substantial input from the experimenter in this process. It is up to the experience
of the experimenter to determine which components should be rejected. There
are, however, some simple statistical techniques that can be used to automatically
select artifactual components. In Chapter 6, a method called ADJUST (Mognon et
al., 2010) is used to suggest ICA components for removal. This method divides
possible artifacts into four classes — three related to eye movements (blinks,
vertical, and horizontal eye movements), and one generic ‘discontinuity’ class. A
class-specific detector is implemented by calculating a set of spatial, temporal, and
joint spatio-temporal features for each component. A threshold is then estimated
for each class, using the entire data-set. Components with features that exceed the
calculated threshold are then deemed to belong to each class of artifact if their
measured temporal, spatial, and spatio-temporal features exceed the threshold (see
Mognon et al., 2010 for full details and validation). In the analysis detailed in
Chapter 6, the list of identified components was used as a ‘shortlist’ of potential
components to reject, and all components were also visually inspected. In most
cases, however, only small changes to the suggested list were made.

4.7 EEG analysis techniques used in this thesis

Figure 4-8 shows an example ERP and the corresponding global field
power (GFP, explained shortly) for data taken from the experiment reported in
Chapter 6. The stimulus was a combined auditory-visual stimulus: a simple visual
flash presented just below a fixation cross on a CRT display, combined with a
short auditory beep. The ERP shows two main waveforms when all electrodes are
viewed together. However, mixed up in these two waveforms will be the auditory
N1 and P2, from temporal, parietal, and frontal sources, as well as the visual P1
and N2, from mostly occipital sources. Depending on the reference location and
the orientation of the different sources, the peaks of all these waves will be
apparent at the scalp in different locations and at different latencies. This
difference in latency between the peak amplitudes measured at each electrode does not necessarily mean that there may be several neural generators active at different times. A single generator with a specific orientation may account for this pattern of activity at the scalp. The ERP to a multi-modal stimulus is thus complex, and it is difficult to measure the peaks of the auditory and visual components separately. For this reason, the ERP analysis in this thesis will not attempt to identify and measure specific ERP components. Rather, a single, reference-free measure of global electrical response strength (the GFP) will be calculated across all electrodes, and compared across conditions. Time periods in which the GFP is significantly different between conditions will thus indicate periods of time in which the overall brain response differs, without regard to specific locations or sources of the difference. In order to make theoretically useful conclusions about these differences, the sources of the generators giving rise to the GFP difference will then be determined by using source localisation techniques, constrained to the time periods at which significant GFP differences occurred.
Figure 4-8. An example of the relation between the average ERP and GFP. The top panel shows a butterfly plot of all 60 channels from a combined audio-visual stimulus consisting of a simple visual flash and auditory beep (data taken from the experiments reported in Chapter 6). The bottom panel shows the GFP calculated from the same data. It would be difficult to determine which peaks or parts of the waveform are related to the auditory and visual stimuli – rather, the whole response is a mixture of the two. As can be seen, the peaks in the GFP correspond with periods of maximum activity in the ERP. However, as the measure is reference-free, no decision about electrode references is needed to find the points of maximum activity.

### 4.8 Global Field Power (GFP)

The global field power (GFP) is a single, reference-free indication of the overall electrical response strength across the whole scalp at a given time (Lehmann and Skrandies, 1980). Formally, it is defined as the sum of squares of
potential differences at all possible electrode combinations at each time point
(Lehmann and Skrandies, 1980, Skrandies, 1989), and can be written as

$$GFP = \sqrt{\frac{\sum_{i=1}^{N}(u_i - \bar{u})^2}{N}}$$  \hspace{1cm} \text{Equation 4-1}

where \(u_i\) is the voltage at electrode \(i\), \(\bar{u}\) is the mean voltage across all
electrodes, and \(N\) is the total number of electrodes. This is mathematically
equivalent to the more straightforward calculation of standard deviation across all
electrodes at each time point (see the Appendix in Murray et al., 2008 for proofs).

Scalp activity across all electrodes that is characterised by pronounced peaks and
steep slopes results in a high GFP value, and the opposite is the case for more
unstructured, ‘flat’ fields at the scalp. Plots of the GFP show periods of high
signal-to-noise ratio, and have been shown to correspond with periods of high

Periods of low GFP are associated with changes in the map configuration.

The advantage of the GFP response is that it is reference-free – the peak of
the GFP response indicates that the overall electric field was strongest at that
particular point in time (in the case illustrated in Figure 4-7, two peaks at around
105 and 220 ms). The disadvantage is that no information at all is provided about
possible locations in the brain that might be generating the fields. In essence, the
GFP provides a reference-free indication of neural response strength at each time
point. It contains no information about how the electric field is distributed across
the electrodes. Once the times of maximum response strength are known,
however, source localisation techniques may be used to determine the possible
sources of the response differences.

### 4.9 Source localisation

Although analysis of the ERP or GFP waveforms can provide precise
indications of the time at which cognitive events of interest occur, only very
approximate indications (or no indication in the case of the GFP) can be made of
the neural sources that give rise to the cognitive events of interest. As has been
mentioned previously, the EEG is generated within the folded-over layers of the cortex. Specific generator sites that are fairly close together may thus be located on the surface of the gyri, just below and parallel with the skull surface, or tangential to the skull surface within a sulcal wall. The scalp topography of these two sources could potentially be radically different, despite their close location within the brain. By taking into account the electrical properties of grey matter, white matter, and the skull and various membrane layers, the translation between neural generators and possible scalp topography can be physiologically modelled, and is often referred to as the ‘forward problem’ (Hallez et al., 2007). The solving of the forward problem results in a transfer function called the ‘lead field,’ and there are variety of lead field solutions of varying complexity and physiological accuracy (Weinstein et al., 2000).

Although the forward problem has been reasonably accurately solved, most researchers have the opposite problem – they are in possession of the scalp-recorded fields as a function of time, and would like to know what the possible neural sources of these fields might have been (the ‘inverse problem’). Unfortunately, the inverse problem is strictly impossible to solve, as an infinite number of possible neural sources, and orientation of sources, could generate any unique field measured at the scalp. The only way of solving the inverse problem is to introduce a number of a priori assumptions, that limit the possible number of solutions. Over the last 10-20 years, many different algorithms and systems of solutions have been proposed. Broadly speaking, these can be classified into those that calculate the most optimal location and orientation of a small number of equivalent current dipoles (over-determined models), and those that calculate three-dimensional current density distributions at every possible three-dimensional location in the brain (under-determined models). Reviews of these various techniques (LORETA, sLORETA, LAURA, MUSIC, FINES, and others) can be found in Michel et al (2004), and Grech et al (2008). Only one technique, sLORETA, will be used in this thesis, and is briefly reviewed here.
4.9.1 sLORETA

Standardised Low-Resolution Electromagnetic Tomographic Analysis (sLORETA) (Pascual-Marqui, 2002, Pascual-Marqui et al., 1994) is an example of an under-determined source model. sLORETA calculates statistical maps of current source density from scalp EEG data that indicate the locations of the underlying sources. The maps are computed by weighting the results from a minimum-norm least squares inverse solution (Hamalainen et al., 1993) by the estimated variance on a voxel-wise basis, with the assumption that the most locally coherent distributions (often referred to as ‘smoothest’) of sources are most likely to result in measureable EEG signals at the scalp. This assumption has been supported by electrophysiological recordings showing highly correlated activity in neighbouring neuronal populations (Haalman and Vaadia, 1997, Llinas, 1988). Given this assumption, the resulting maps are more spatially diffuse than is likely to be the case (Luck, 2005), however the centre of the resulting activations still provides an accurate representation of the likely generator of the input field.

sLORETA produces estimates of the 3D current sources of the scalp EEG in 6239 cortical gray matter voxels, with a resolution of 5 mm according to the digitized Talairach (Lancaster et al., 2000) and probability atlases of the Brain Imaging Center, Montreal Neurological Institute (Mazziotta et al., 2001). The head model for this inverse solution uses the electric potential lead field computed with the boundary element method applied to the MNI152 template (Fuchs et al., 2002). Electrode coordinates on the MNI152 scalp for the 10/5 system are based on a head-surface-based positioning system (Jurcak et al., 2007).

4.10 Conclusion

This chapter has provided an overview of some of the methodologies that are used in sections of the experimental work in Chapters 5-7. In Chapter 6, the EEG analysis will use these techniques in a four-stage analysis.

1) Pre-processing using ICA for artifact correction.
2) Viewing of multisensory difference waves across all electrodes with the average reference.

3) Point-wise significance testing of the GFP in order to determine time intervals in which a significant difference between uni- and multi-modal stimuli occurred.

4) Source localisation during the intervals where significant differences were found.
5 Does spatial incongruity affect the flash-beep illusion?


Reproduced exactly as published, apart from figure/heading labels and other minor formatting changes as required. Copyright remains with the authors.
5.1 Introduction

In cases where multiple senses provide congruent information concerning the same external event, multisensory integration can result in perceptual advantages. For instance, the detection of weak visual stimuli is enhanced with a concurrent auditory stimulus (Frassinetti et al., 2002a), and speech perception is improved when video of the moving lips is available (Sumby and Pollack, 1954, MacLeod and Summerfield, 1987). However, if an experimenter arranges incongruent information to be presented to each sensory modality, the senses can also interfere with each other, causing altered or illusory percepts. Perhaps the best-known example of this phenomenon is the McGurk-MacDonald illusion (McGurk and MacDonald, 1976, MacDonald and McGurk, 1978), where vision alters the way a speech sound is perceived. When the audio for “ba” is presented along with video of an actor pronouncing “ga”, the consonant half-way between (“da”) is perceived. In the majority of these illusions, visual information tends to dominate in the spatial domain.

In rarer cases, particularly where temporal information is involved, audition may dominate visual perception. The sound-induced flash illusion (Shams et al., 2000) is one such case. In this experiment, a varying number of flashes were presented along with a varying number of short beeps. When a single flash was presented along with two or more beeps, observers often reported seeing two or more flashes. In an event-related potential (ERP) study of the same illusion (Shams et al., 2001), the auditory stimulus was found to modify the flash visual evoked potential (VEP). The authors proposed that the illusion is perceptual – that the second auditory stimulus caused the illusory perception of a second flash. Other evidence from EEG (Bhattacharya et al., 2002, Mishra et al., 2007) and MEG (Shams et al., 2005a) has shown neurophysiological correlates of the illusion in primary visual areas, thus giving weight to the suggestion that the illusion is perceptual rather than a result of response biases. fMRI studies (Watkins et al., 2006, Watkins et al., 2007, Zhang and Chen, 2006) have also found increased functional activity associated with the illusion in V1 as well as the superior
temporal sulcus (STS) and superior colliculus (SC) (Watkins et al., 2006). Both these structures have previously been associated with the integration of auditory and visual information (Jiang et al., 2002, Beauchamp, 2005, Calvert, 2001). A corresponding flash-fusion illusion has also been reported, where a single beep causes fusion of a double flash stimulus (Andersen et al., 2004, Mishra et al., 2008). Few other behavioural investigations of the illusion have been performed (although see McCormick & Mamassian (2008) for a recent exception).

Studies using single-cell electrophysiology (Stein and Meredith, 1993) have found that certain neurons in the superior colliculus (SC) are responsive to stimulation in more than one sensory modality. Interestingly, many of these cells show a response gradient where visual and auditory stimuli occurring in close spatial and temporal proximity cause cells to respond more strongly than would be expected by simple addition of the individual responses to each uni-sensory stimulus. These properties have been described by three rules (Stein and Meredith, 1993). The spatial rule states that only stimuli from different modalities that are in close spatial proximity are integrated and produce response enhancement. This reflects the manner in which maps of space across different modalities are aligned in the SC. The temporal rule states that multisensory stimuli are more likely to be integrated when they occur at similar times, reflecting the way in which maximal multisensory enhancement occurs when the responses to each unisensory stimuli are at their peaks (Meredith et al., 1987). The inverse effectiveness rule describes how individually weak unisensory stimuli combine to produce larger neural responses than would be expected by the simple addition of each individual response. Together, these rules have provided a conceptual framework for mapping the behavioural consequences of multisensory integration to possible underlying physiological properties. These physiological data seem to correlate well with human behavioural performance, which also shows spatial (Bolognini et al., 2005b) and temporal (Frassinetti et al., 2002a) response gradients. The sound-induced flash illusion is known to gradually vanish with increasing temporal incongruence (Shams et al., 2002), and thus is consistent with the temporal rule, however little is known about how spatial incongruence might affect the illusion.
In addition to structures such as the SC, multisensory integration can also occur rapidly via direct cortico-cortical pathways (see Driver and Noesselt (2008) for a review). Direct projections to V1 from the core and belt regions of the auditory cortex as well as upper banks of the superior temporal sulcus (STS) have been found in monkeys (Falchier et al., 2002), and these connections appear to target mainly the peripheral region of the visual field (Falchier et al., 2002). Somatosensory-responsive regions within human auditory cortex have been found using fMRI (Foxe et al., 2002) however these early interactions do not appear sensitive to spatial congruence (Murray et al., 2005).

In the environment, auditory and visual signals from a common event or object are often spatially and temporally congruent, and the mechanisms underlying multisensory integration seem bound by corresponding rules, which are in turn reflective of the underlying properties of multisensory neurons. The aim of this study is to investigate the effect of spatial incongruence in the sound-induced flash illusion. If the illusion occurs less strongly when auditory and visual stimuli are separated, then integration supporting the illusion must be strongly dependent on spatial congruence. In this case the illusion would be consistent with both the spatial and temporal rules describing response properties of multisensory neurons in the SC. Conversely, if perception of the illusion is unaffected, and multisensory integration occurs regardless of the spatial congruence of the auditory and visual stimuli, then the “spatial rule” may not apply to this illusion, suggesting that multisensory integration could be subserved by parts of the brain not known to be governed by these rules.

5.2 Methods

5.2.1 Ethics Statement

The study conforms to The Code of Ethics of the World Medical Association (Declaration of Helsinki), and was approved by the Swinburne University Human Research Ethics Committee. All participants gave informed consent.
5.2.2 Participants

Nine healthy adults (five male, $M=26.9$ years, $SD=4.2$) participated in the study after providing written informed consent. Each participant reported normal or corrected-to-normal vision and hearing.

5.2.3 Stimuli

In experiment one, the stimulus configuration followed the original report (Shams et al., 2002) as closely as possible. The experiment was conducted in a quiet room with an average background sound level of 39 dB (A-weighted). Brief flashes were presented on a cathode-ray tube (CRT) computer monitor. Short beeps were presented along with the flashes from two small speakers placed centrally under the monitor. The centre of the speaker was 8 degrees below the visual stimulus.

The visual stimulus consisted of a white disk on a black background flashed from one to four times. The disk subtended 2° of visual field and was located 5° below a small central fixation cross. The refresh rate of the CRT monitor was set to 60 Hz (the refresh period was thus 16.7 ms), and each flash was set to display for one refresh period of the monitor. In order to determine the exact flash duration the persistence of the CRT phosphors was measured using a photodiode and oscilloscope. The flash duration was 1.3 ms (full width at half maximum - FWHM). In trials where more than one flash was presented, the next flashes followed after three blank refresh periods. The between-flash onset asynchrony was thus 67 ms.

The beep was a 3500 Hz, 85 dB (A-weighted) sine wave of 8 ms duration (3 ms rise/fall times). On one-flash trials, zero to four beeps were presented. On multi-flash trials, either zero or one beep was presented. For relative timing information see Figure 5-1. The first beep was always presented 23 ms prior to the first flash, and in trials where more than one beep was presented, the between-beep asynchrony was 57 ms. The stimulus combinations will be referred to using
abbreviations for the number of flashes (F) followed by the number of beeps (B), such that “1F0B” refers to a single flash with no beeps.

In the second experiment, the basic physical characteristics of the stimuli were identical to those used in experiment one. In experiment two however, the flashes could be presented 10° either to the right or left side of the fixation cross. The beeps were presented from two small speakers, each placed immediately below the screen at the eccentricity that the flashes were to appear. The 20° separation thus obtained is comfortably larger than the minimum discernable angles for auditory (Perrott and Saberi, 1990) and visual stimuli. Although simple sinusoid stimuli were used that may be hard to locate due to their reduced spectral complexity, they were presented in free-field conditions designed to maximise ease of localization.

Figure 5-1. Stimulus timing for flashes and beeps. A 2-flash 2-beep (2F2B) trial is shown here.

5.2.1 Procedure

In the first experiment, participants sat in a chair with their eyes at a distance of 70 cm from a computer screen. The fixation cross was displayed alone for an interval that varied randomly on each trial between 1200 and 1500 ms. The flash/beep sequence then began. Following the sequence was another short randomly varied interval (1200 to 1500 ms), after which the text “How many flashes did you see?” was displayed in place of the fixation cross. This text
remained in place until the participant made a response on the keyboard. Participants were instructed to keep their gaze on the fixation cross during each trial and count the number of flashes that would appear whilst ignoring the beeping sounds. The response was made after each trial by pressing keys numbered one to four on a keyboard. Each of the eleven possible beep/flash combinations (1F0B, 1F1B, 1F2B, 1F3B, 1F4B, 2F0B, 2F1B, 3F0B, 3F1B, 4F0B, 4F1B) was presented randomly in a single block. This block was repeated five times (with trials re-randomised each time).

In the second experiment, the sources of the auditory and visual stimuli were separated. As in experiment one, during one-flash trials zero to four beeps were presented. On multi-flash trials, either zero or one beep was presented. There were four possible spatial configurations: two congruent configurations where the beep and flash are presented together on the left and together on the right (termed LL and RR), and two incongruent configurations where flashes on the right are paired with beeps on the left, and vice versa (termed LR and RL). On the seven trials when both beeps and flashes were present (1F1B, 1F2B, 1F3B, 1F4B, 2F1B, 3F1B, 4F1B), all four spatial configurations were displayed (LL, LR, RL and RR), and for those with only flashes present (1F0B, 2F0B, 3F0B, 4F0B), only two configurations were possible (LL and RR). There were thus 36 types of trials. The experiment was broken into five blocks, in which each trial type was presented five times. Within each block trials were ordered randomly.

5.3 Results

5.3.1 Central presentation - mean responses

Figure 5-2 (panel A) shows the mean responses averaged across the nine participants (error bars show standard error of the mean) when a single flash was presented. Increasing the number of accompanying beeps dramatically increased the number of flashes reported when only one flash was present. The increase was strongest in the case where one flash was accompanied by two beeps compared to a single beep. Increasing the number of accompanying beeps to three increased
the average response further, but the effects were not significant. In order to test the effect of beeps on the number of flashes reported, the mean responses for the five 1-flash trials were submitted to a repeated-measures Analysis of Variance (ANOVA), with a 5-level within-subjects factor Beep (0-4 beeps). The main effect of Beep was significant ($F(4,32) = 17.5, p < 0.01$). When a single flash was presented the number of beeps had an effect on the number of flashes reported. Post-hoc comparisons revealed a significant increase in the mean number of flashes reported from 1.2 ($SD = 0.3$) when one beep was presented to 1.8 ($SD = 0.2, p = 0.001$) when two beeps were presented.

![Figure 5-2](image)

**Figure 5-2.** Mean responses for the number of flashes reported (bars show SEM) for all trial types in experiment one. Single-flash trials are shown in panel A, and multi-flash trials in panel B. The reference line in panel B shows veridical perception.

Figure 5-2 (panel B) shows trials where more than one flash was presented, with separate lines for trials with zero and one beep. It can be seen that there is very little difference between trials with or without beeps, indicating that the presence of a single beep did not affect the ability of the participants to correctly count any number of flashes. To test whether responses increased as the number of flashes presented increased, and whether this relationship changed depending on whether beeps were presented concurrently, the mean responses were
submitted to a repeated measures ANOVA. There were two within-subjects factors – Flashes (1, 2, 3 or 4) and Beeps (0 or 1). There was a significant main effect of Flashes \( F(3,24) = 176.8, p < 0.001 \) but no significant effect of the number of beeps. Post-hoc tests indicated that the number of flashes reported increased significantly between each adjacent level of flashes presented.

The reference line in Figure 5-2 (panel B) indicates veridical perception – the responses that could be expected if flash perception were perfectly accurate. It can be seen that for one to three flashes, the measured responses (whether a beep was presented concurrently or not) are close to the reference line, indicating that for up to three flashes the stimuli are relatively easy to perceive. For four flashes however, participants consistently under-report the number of flashes presented \( (M = 3.2, SD = .1) \). Further analysis will therefore exclude the four-flash stimuli on the basis that they are ambiguous.

After excluding the four-flash stimulus, a linear regression analysis showed that the number of flashes reported was a highly significant predictor of the number of flashes actually presented \( (R^2 = 0.86, F(1,52) = 367.4, p < 0.001) \), with a nearly unitary slope \( (\beta = 0.93, t(52) = 19.2, p < 0.001) \), indicating that as the number of flashes presented increased by one, the number flashes reported was also likely to increase by one.

### 5.3.2 Central presentation - response rates

Following (Andersen et al., 2004), the stimuli were divided into three classes – those capable of producing fission illusions (1F2B, 1F3B), those capable of producing fusion illusions (2F1B, 3F1B) and the unimodal visual stimuli (1F0B, 2F0B, 3F0B). A categorical analysis was then performed on the response counts for fission and fusion stimuli. Mean response counts (expressed as a percentage of the total number of trials) for each stimulus type are shown in Figure 5-3 (wide grey bars only).
Counts of fission responses were calculated by summing the number of responses indicating more flashes than were presented for the two fission-capable stimuli. Similarly, non-fission counts were computed by summing responses indicating the correct number or fewer flashes than were presented. The ratio of fissions:non-fissions then indicated the odds of a fission illusion occurring with a multisensory stimulus. For both fission stimuli the odds of fission with the equivalent unisensory stimulus (1F0B for both fission stimuli) was similarly determined. Finally, the odds ratio of fission in the multisensory case vs. the unisensory equivalent was calculated. A value of greater than one in this final odds ratio thus indicates a greater likelihood of fissions for the audiovisual stimulus compared to its unimodal equivalent. The significance of the association between the stimulus type (uni vs. multi-sensory) and the report of illusions (fission/fusion vs. non-fission/non-fusion) was assessed with a Chi-square test. To follow (Andersen et al., 2004), the significance of the association was also tested using...
Fisher’s exact (one-sided) test, although it should be noted that two out of the four cells in the fusion Chi-square test have expected counts of less than five, violating the assumptions for that test. The results of both these tests, along with the corresponding odds ratios, are shown in Table 5-1. Both fission stimuli showed a significant association between the type of stimulus (uni- or multi-sensory) and the report of an illusion, with both multisensory stimuli approximately 30 times more likely to elicit an illusion response than unisensory stimuli. Neither fusion stimulus showed a significant association.

<table>
<thead>
<tr>
<th>Stimulus</th>
<th>Odds of illusion (multisensory)</th>
<th>Odds of illusion (unisensory)</th>
<th>Odds ratio</th>
<th>$\chi^2$ (df)</th>
<th>$\chi^2$ p</th>
<th>Fisher $p$ (one-sided)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1F2B (fission)</td>
<td>3.1</td>
<td>.25</td>
<td>12.4</td>
<td>27.8(1)</td>
<td>&lt;.000</td>
<td>&lt;.000</td>
</tr>
<tr>
<td>1F3B (fission)</td>
<td>4.6</td>
<td>.25</td>
<td>18.5</td>
<td>34.9(1)</td>
<td>&lt;.000</td>
<td>&lt;.000</td>
</tr>
<tr>
<td>2F1B (fusion)</td>
<td>.02</td>
<td>.02</td>
<td>1.0</td>
<td>1.0(1)</td>
<td>.3</td>
<td>.5</td>
</tr>
<tr>
<td>3F1B (fusion)</td>
<td>.03</td>
<td>.02</td>
<td>2.6</td>
<td>2.7(1)</td>
<td>.09</td>
<td>.08</td>
</tr>
</tbody>
</table>

Table 5-1. The odds for fusion and fission illusions for each of the four illusion-capable stimuli in Experiment One (central presentation). The final odds ratio, and the significance of the association between stimulus type (uni- vs. multi-sensory) and reports of illusions (fissions/fusions vs non-illusory) are also shown.

### 5.3.3 Spatial presentation - mean responses

As in experiment one, mean responses were calculated for each trial type, and are displayed in Figure 5-4 (panel A). To determine the effect of the number of beeps in the four different spatial configurations on the number of flashes reported, the mean responses for the sixteen stimulus types in which both beeps and flashes were present were submitted to a 4 (Beeps: 1, 2, 3, 4) x 4 (Stimside: LL, LR, RL, RR) repeated measures ANOVA. There was again an effect of Beeps ($F(3,24)= 34.0, p < 0.001$), but no significant effect of Stimside, and no interaction between the number of beeps and the Stimside condition. As in experiment one, contrasts revealed that the number of flashes reported increased significantly when
the number of beeps presented increased from one to two ($F(1,8) = 114, p < 0.001$).

To investigate whether participants showed any spatial bias towards either the left or right in their ability to count the flashes in the absence of any auditory stimulus, the mean responses when no beeps were present were examined Figure 5-4 (panel B). For the eight stimulus types in which flashes were presented alone, mean responses were submitted to a 2 (Stimside: LL, RR) x 4 (Flashes: 1, 2, 3, and 4) repeated measures ANOVA. There was a significant main effect of the number of flashes presented ($F(3,24) = 298, p < 0.001$), but no effect of the side that the flash was presented ($F(1,8) = 0.001, p = 0.9$).

Next, the single-beep trials were analysed Figure 5-4 (panel C). Analysis of these trials allowed examination firstly of whether separation of the auditory and visual components of the stimulus affected the participants ability to count flashes overall (similarly to the zero-beep trials), and secondly of whether there might be a change in this ability depending on the exact stimulus configuration. Mean responses for single-beep trials were submitted to a 4 (Flashes: 1, 2, 3, and 4) x 4 (Stimside: LL, LR, RL, RR) repeated-measures ANOVA. There was a main effect of the number of flashes ($F(3,24) = 436, p < 0.001$), but no main effect of the stimulus configuration, nor an interaction between the stimulus configuration and the number of flashes.

![Figure 5-4. Mean responses for the number of flashes reported (bars show SEM) for all trial types in experiment two. Single-flash trials are shown in panel A. Multi-flash trials with zero beeps and one beep are shown in panels B and C respectively.](image-url)
5.3.4 Spatial Presentation - Response rates

As the stimulus configuration was found to have no effect on responses to the flash stimulus in illusion conditions or otherwise, the stimulus configuration factor was collapsed and the data from experiments one and two were pooled. However, as trials with both flashes and beeps had four possible spatial configurations compared with the two possible with flash-only trials, only the LL and RR trials (common to all trial types) were added. The response rates using this larger number of trials were then re-analysed. To enable direct comparison with results from (Andersen et al., 2004), the odds ratio calculated for the sums of all fission stimuli and all fusion stimuli were computed. The results are shown in Table 5-2. The pattern of results is similar to those from Experiment 1. Both fission stimuli show significant associations between the stimulus type and the odds of an illusion response, while both fusion stimuli show no significant association. For the summed fusion and fission stimuli the pattern is again the same – the association between stimulus type and the odds of an illusion response is significant for fission stimuli, but not fusion stimuli.

<table>
<thead>
<tr>
<th>Stimulus</th>
<th>Odds of illusion (multisensory)</th>
<th>Odds of illusion (unisensory)</th>
<th>Odds ratio</th>
<th>$\chi^2$ (df)</th>
<th>$\chi^2$ p</th>
<th>Fisher p (one-sided)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1F2B (fission)</td>
<td>.4</td>
<td>.16</td>
<td>24.4</td>
<td>17.8(1)</td>
<td>&lt;.001</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>1F3B (fission)</td>
<td>.38</td>
<td>.16</td>
<td>22.34</td>
<td>115.2(1)</td>
<td>&lt;.001</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>2F1B (fusion)</td>
<td>.04</td>
<td>.02</td>
<td>2.0</td>
<td>1.0(1)</td>
<td>.3</td>
<td>.5</td>
</tr>
<tr>
<td>3F1B (fusion)</td>
<td>.16</td>
<td>.19</td>
<td>.8</td>
<td>.26(1)</td>
<td>.6</td>
<td>.7</td>
</tr>
<tr>
<td>All fission</td>
<td>3.9</td>
<td>.10</td>
<td>38.3</td>
<td>233.0(1)</td>
<td>&lt;.001</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>All fusion</td>
<td>.10</td>
<td>.10</td>
<td>1</td>
<td>0(1)</td>
<td>1.0</td>
<td>1.0</td>
</tr>
</tbody>
</table>

Table 5-2. The odds of fusion and fission illusions for multisensory stimuli and their uni-sensory equivalents in combined data from experiments 1 and 2. The LL and RR spatial configurations were collapsed and pooled with the data from experiment one. The rows labeled “All fission” and “All fusion” show data summed across all fission- and fusion-capable stimuli, respectively. The final odds ratio, and the significance of the association between stimulus type (uni- vs. multi-sensory) and reports of illusions (fissions/fusions vs non-illusory) are also shown.
5.4 Discussion

The experiments reported here replicate the flash-beep fission illusion described by Shams et al. (2002, 2000), however we find no evidence for an associated flash-beep fusion illusion reported in later studies (Andersen et al., 2004, Mishra et al., 2007, Watkins et al., 2007, Shams et al., 2005b). We also show that the flash-beep illusion is not affected by spatially separating the auditory and visual components of the stimulus. These results suggest that although the illusion has been previously shown to be compatible with the temporal rule describing neural activation in the SC, it is not bound by the corresponding spatial rule. The illusion thus may have neural origins that are outside the regions governed by these rules.

The results in experiment one were similar to those obtained by Shams et al. (2002) – when two or more beeps were presented along with a single flash, participants reported seeing two or more flashes. The number of flashes reported increased from approximately one, to approximately two when the number of beeps presented increased from one to two. As in the Shams et al. study, the number of flashes reported did not increase significantly beyond two as the number of beeps increased further to three and four.

The results in the control trials also echoed those in the Shams et al. study. In zero-beep trials, it was found that participants were able to count up to three flashes relatively easily, indicating that the visual stimulus when presented alone was unambiguous up to that point. However, in the Shams et al. study the participants appear to be more accurate with four flashes than in the present study (compare their Fig. 4A with our Fig. 2B). In the present study, there were no differences in the number of flashes reported between zero-beep and one-beep trials. This held for any number of flashes presented, and indicates that the presence of an auditory stimulus per se does not appear to affect the number of flashes reported. After excluding the four-flash trials and pooling the zero-beep and one-beep trials, the number of flashes presented was a significant predictor of the number of flashes reported, again indicating that our participants had no
difficulty reporting the number of flashes presented. In summary, with the exception of the four-flash stimuli, the results in the first experiment closely followed those reported by Shams et al. (Shams et al., 2002), suggesting that any differences in stimuli, participants or procedure did not lead to a reduction in the strength of the illusion.

In their study, Shams et al. (Shams et al., 2002) found that an illusion occurred only when the number of beeps exceeded the number of flashes and not vice versa. The well established “modality appropriateness” hypothesis (Welch and Warren, 1980) thus could not explain their results. Audition provides more accurate temporal information than vision, and thus should be the more appropriate modality for this task. According to this theory, the number of beeps should thus always dominate over the number of flashes. However, in the results from Shams et al. (Shams et al., 2002), this did not appear to be the case, and the authors proposed instead the “discontinuity hypothesis” – that the discontinuous stimulus in one modality alters the percept of the continuous stimulus in the other modality. The present results fit with this pattern, thus also suggesting that the modality appropriateness hypothesis does not fully explain the flash-beep illusion.

In addition to the sound-induced flash (or fission) illusion, Andersen et al. (Andersen et al., 2004) were the first to report a corresponding sound-induced fusion illusion. In their study, participants reported seeing fewer flashes than were actually presented on trials with fewer beeps than flashes. This fusion effect was weaker than the fission illusion, and disappeared when the auditory stimulus was reduced to a near-threshold level. Rather than supporting one theory or the other, the authors suggested that modality appropriateness and discontinuity are both factors that combine to influence the dominance of each modality. Several other studies have now shown similar results (Mishra et al., 2008, Shams et al., 2005b, Watkins et al., 2007), with participants reporting fewer flashes than were presented in trials with fewer beeps than flashes. In order to compare the present results with Andersen et al. (2004) our data were re-analysed following their categorical analysis technique. The results are strikingly different – firstly the likelihood of fission
illusions was almost five times higher in the present data (see Table 2), and secondly no evidence was found for fusion illusions occurring.

There are several possibilities for the discrepancy in these results. Firstly, in the present study the stimulus timing in Shams et al. (2002) was followed as closely as possible, with the auditory stimulus leading the visual by 23 ms (when both auditory and visual stimuli were present). Andersen et al. (2004) used a simultaneous presentation of the flash and beep. It is possible that by presenting the auditory stimulus 23 ms before the flash, multiple flashes were somehow rendered less likely to fuse than if they were presented simultaneously. In other respects the stimuli are closely matched. Other studies that have reported a fusion illusion in the 2F1B condition (Mishra et al., 2007, Watkins et al., 2007) have also presented the auditory and visual stimuli simultaneously. However one study using simultaneous auditory-visual stimuli (Meylan and Murray, 2007) did not find any evidence of fusion, and another (Shams et al., 2005b) used non-simultaneous auditory and visual stimuli, and did find auditory-visual fusion. The only study to directly investigate the effect of stimulus timing on the illusion (Shams et al., 2002) found that varying the auditory-visual onset by ±70 ms had little effect on the strength of the fission illusion, but with separation beyond 70 ms the strength of the illusion gradually declined. However, only the fission-illusion stimuli were included in their investigation. Stimulus timing may have a different effect on different stimulus combinations.

Another influence on stimulus timing is the nature of the display system itself. The flash stimulus is typically quoted as being of 17 ms duration, with a flash onset asynchrony of 67 ms (with the exceptions of Meylan & Murray (2007) – duration 13ms, onset asynchrony 65 ms; Mishra et al (2007) – duration 5 ms, onset asynchrony 70 ms, and Shams et al. (2005b) – duration 10 ms, onset asynchrony 70 ms). The flash durations reported in the literature generally correspond to the refresh period of the display (usually set at 60 Hz, giving an intended “flash duration” of 16.7 ms) rather than the measured duration of the flash. However, different display types can have widely varying decay times, so that a single refresh of a liquid crystal display (LCD) monitor or projector for instance may produce a
flash stimulus substantially longer than that produced by a CRT monitor, despite the refresh rates and intended flash durations being identical. In the present study, the exact duration of a single-refresh flash of the CRT set to a refresh rate of 60 Hz was measured using a photodiode and oscilloscope at 1.3 ms (FWHM), rather than the 17 ms that would be quoted by assuming that the flash duration is equal to the refresh rate period. An examination of a variety of other CRT and LCD monitors revealed similar flash durations among CRT monitors (1.2 – 1.5 ms FWHM), but much longer flash durations for LCD screens (12 ms FWHM). A longer flash stimulus would result in a shorter inter-flash interval in multi-flash trials, and may render two consecutive flash stimuli more likely to fuse. Of the four studies finding evidence for a fusion illusion, two have used LCD projectors (Watkins et al., 2007, Watkins et al., 2006), made necessary as the experiments were performed in a functional magnetic resonance imaging scanner. A light-emitting diode was used in another (Mishra et al., 2007), and Andersen et al. (2004) and Shams et al. (2005b) do not report the display used.

Secondly, while Andersen et al. (2004) used all twelve possible combinations of stimuli (0-3 beeps and 1-3 flashes), a reduced set of eight stimuli was used in the present study. As a result, counts for the frequency of fission and fusion illusions were pooled across two stimulus types (fissions: 1F2B + 1F3B, fusions: 2F1B + 3F1B) rather than the three used in Andersen et al. (2004). For fusion stimuli in particular this may have had a large effect. In Andersen et al (2004), only 50% of responses to the three-flash stimulus were correct even in the absence of any auditory stimulus. It is therefore difficult to interpret any further effects of the auditory stimuli in that study. In the present study, participants were more accurate with the 3F0B (80% correct) and 3F1B (75% correct) stimuli (Fig. 3), again possibly due to differences in the display characteristics. With the visual stimulus more easily perceived, it might be that any fusing effect of the auditory stimulus was rendered less effective in the present study.

After validating the procedure, the effect of spatially separating the auditory and visual components of the multisensory stimuli in the illusion was examined. The flash-beep illusion can so far be understood in terms of only one of
the three multisensory rules – the temporal rule. The illusion is strong with temporal variability up to around 70 ms, after which point it gradually disappears (Shams et al., 2002). This figure ties in closely with temporal integration times for multisensory neurons in the superior colliculus (Meredith et al., 1987). Shams et al. (Shams et al., 2002) theorize that the illusion is the result of auditory processing modifying visual perception rather than a result of decision-making biases. If true, it is thus likely to at least in part be supported by functions of the SC, and is already known to obey the temporal rule. However, the present results indicate that despite separating the stimuli in space to an extent that should be easily perceived, no aspect of performance on the task changed, both on illusion and control trials.

There have now been many electrophysiological studies in cats and monkeys (Stein and Meredith, 1993, Stein et al., 1993) indicating the existence of neurons in the SC that respond in an integrative manner to multisensory stimuli. These neurons are sensitive to temporal and spatial congruence between the auditory and visual signals from a multisensory stimulus – as the two signals are moved further apart in time or space, the responses from these multisensory neurons reduce. This phenomenon has been echoed at a behavioural level in humans, with performance on multisensory tasks exhibiting similar spatial (Bolognini et al., 2005b) and temporal (Frassinetti et al., 2002a) response gradients. In (Frassinetti et al., 2002a), participants were asked to fixate centrally and detect sub-threshold masked flash stimuli displayed at 8, 24, 40 and 56 degrees in the left and right visual fields. The task was performed in a vision-only condition, as well as an auditory-visual condition where sounds were presented either at the same or different locations to the flash. They found that perceptual sensitivity generally improved when an auditory signal was presented at the same location. Interestingly, there was one exception to this finding: when an auditory stimulus was presented 16º further to the right of the visual stimulus at 40º, perceptual sensitivity was enhanced. When the sound was played at other locations there was no improvement, despite adjacent locations being only 16º apart. These results are thought to reflect the fact that performance on such tasks is closely related to
multisensory processes in the SC. In a follow-up study (Maravita et al., 2008), a simple audio-visual detection task was administered using red or blue/purple visual stimuli. Blue/purple stimuli are detected using only S-type cones in the retina, which do not project directly to the SC (Schiller and Malpeli, 1977, de Monasterio, 1978). Reaction time measures showed evidence for multisensory integration only using the red stimuli (which do project to the SC), providing evidence for the involvement of the SC in audiovisual multisensory integration in humans. Furthermore, the reaction time effect for red stimuli diminished when the sources of the auditory and visual stimuli were spatially separated, and when a 250 ms delay was added between the auditory and visual stimuli. In comparison, although the sound-induced flash illusion has been shown to be consistent with the temporal rule, in the present experiment separating the spatial origins of the visual and auditory stimulus by 20º did not change reports of the illusion in any detectable way. The results suggest that the rules of multisensory integration as they apply to neurons in the SC may not hold in the case of the flash-beep illusion.

While most studies find that the facilitatory effect of multisensory stimulation requires that the two stimuli be spatially coincident (Harrington and Peck, 1998, Frens et al., 1995, Frassinetti et al., 2002a), this does not always seem to be the case. In a series of three experiments Stein and colleagues (Stein et al., 1996) found that during a luminance intensity judgment task, participants reported higher intensities on trials in which there was an accompanying irrelevant auditory stimulus, even when the sound was located at a random position up to 45 deg away from the visual stimulus. However, when both the auditory and visual stimuli were moved away from fixation, luminance judgements were no longer enhanced. This result is contrary to many that fit with the spatial rule describing properties of multisensory neurons in the SC (Meredith and Stein, 1986, Wallace and Stein, 1994) as well as behaviourally in humans (Bolognini et al., 2005b, Driver and Spence, 1998, Frassinetti et al., 2002b). The authors note that although the spatial rule is well established at the level of individual neurons in the SC, there are multisensory neurons located in many other areas of the brain involved in many different tasks for which stimulus localisation in space is possibly not vital. They
suggest that the task of assessing stimulus intensity could be one such activity. The results from the current study follow a similar pattern – behaviour related to the illusory phenomenon seems to echo known temporal rules for multisensory integration in the SC, but not spatial rules.

If the flash-beep illusion is indeed a case of a visual percept induced by an auditory stimulus, the question remains as to how in the brain this might occur, particularly in the timeframe available. It is becoming increasingly recognised that multisensory integration can take place not only via the SC and other traditionally multisensory areas of the brain, but also directly between primary sensory cortices (see (2008, Ghazanfar and Schroeder, 2006) for reviews). This view is in agreement with findings indicating that sensory-specific judgements (such as contrast for vision) can be affected by information from another sense. In the context of the sound-induced flash illusion, this has recently been seen in a study indicating that the illusory flash has a detectable contrast (McCormick and Mamassian, 2008). Using event-related potentials, it has also been shown that while late (more than 100 ms) multisensory effects on visual processing are sensitive to the spatial congruence of the stimuli (McDonald et al., 2003), the earliest (~50 ms) multisensory effects generally occur irrespective of the stimulus location (Murray et al., 2005), but are sensitive to temporal congruence (Senkowski et al., 2007). The results from the present study generally fit with this view – in order for the auditory stimulus to affect visual perception, integration of the auditory information would have to occur very rapidly, perhaps via direct cortical pathways that are not sensitive to spatial information.

It also worth pointing out that retrograde tracing studies in primates have found that the inputs from the core and parabelt auditory cortex to primary visual cortex provide their strongest connections to areas of primary visual cortex that subserve the peripheral visual field – from 10-20º eccentricity (Falchier et al., 2002). The sound-induced flash illusion is strongest when the flash is presented in the periphery, and Shams et al (Shams et al., 2001) have shown that illusion-related modulations of the flash VEP occur only when the flashes are presented in the visual periphery. Generally an eccentricity of 5-20º has been used in studies of the
illusion, with most studies presenting stimuli within the range where connections from primary auditory to primary visual cortex have been found subserving the visual periphery (Falchier et al., 2002). Although the minimum audible angle for sounds is only a few degrees, spatial receptive fields of individual neurons in the auditory cortex are large, generally occupying a quadrant or more of acoustic space with stimuli 10-30 dB above neural thresholds (Brugge et al., 2001). For a recent explanation of how individually broad receptive fields are thought to translate to high spatial accuracy, see (Miller and Recanzone, 2009). With higher sound levels, the receptive fields broaden still further (Mrsic-Flogel et al., 2005), despite higher-intensity sounds being easier to localise (Altshuler and Comalli, 1975). These direct cortico-cortical connections from neurons in the auditory cortex with very wide spatial receptive fields to neurons subserving the periphery in primary visual cortex is an alternative cortical mechanism that may underlie the illusion, and its insensitivity to spatial incongruence.

In summary, the present results firstly replicate the sound-induced fission illusion described by Shams et al. (2002, 2000). In our replication however we do not find an associated flash-beep fusion illusion (Andersen et al., 2004), and speculate that the variety of findings in the literature concerning the fusion illusion may be due to unintended differences in stimulus timings brought about by different visual displays. Secondly, the present results show that the illusion is insensitive to spatial incongruence of the auditory and visual stimuli. Although the design of the present study cannot directly address the nature of the underlying neural mechanisms, the pattern of results are consistent with recent research (Driver and Noesselt, 2008, Murray et al., 2005, Senkowski et al., 2007) showing that direct cortical connections between primary sensory areas (Falchier et al., 2002) may be sensitive to temporal, but not spatial congruence.
6 the effect of UNI vs MULTI-modal context on visual processing
6.1 Introduction

The experiments described in the previous chapter showed that the flash-beep illusion could be reproduced successfully in our lab, and that it was insensitive to separation of the auditory and visual components of the multi-modal stimulus by 20 degrees of arc. The original aim of the study in the current chapter was to compare neurophysiological responses to illusion-capable stimuli, depending on whether or not an illusion was reported. As will be further explained in the methods section, it was difficult to directly compare these ERP responses. Despite efforts to minimise the number of trials that were rejected due to artefacts, only four of the fourteen participants had accuracy scores for illusory stimuli that provided enough trials to analyse ERPs to both correct and incorrect responses. However, it was still possible to explore possible underlying causes of the illusion, by isolating neural responses to the second flash, depending on whether it was preceded by a uni-modal or multi-modal stimulus. A change in response to the second flash depending on the immediately prior context thus indicated a differential and lasting effect on visual processing, dependant on the presence of a multi-modal stimulus. Although the fission illusion stimuli were not directly tested, the neural mechanisms that allow this kind effect on visual processing may be similar to those underlying the flash-beep illusion.

In this study, 1 or 2 flashes were paired with 0, 1 or 2 beeps, and participants were asked to indicate the number of flashes that they saw. The methods, procedure and apparatus used were similar to those described in Chapter 5, however all stimuli were presented centrally rather than laterally, and several changes were made to the procedure in order to allow EEG recording and analysis. Briefly, these changes involved increasing the number of repetitions of each trial type, and correspondingly reducing the number of different trial types in order to maintain a reasonable experiment length for participants. In order to simplify interpretation of the data and bring the procedure in to line with an increasing number of flash-beep illusion studies in the literature, the timing of the
stimuli was also changed so that auditory and visual stimuli were presented simultaneously rather than with a short delay.

In order to examine the response to the second flash, depending on its uni-modal or multi-modal immediate prior context, two difference waves were calculated. Subtracting the 1F0B response from the 2F0B response isolated neural responses to the second flash with a uni-modal context (when it was immediately preceded by another flash). Likewise, subtracting responses to the 1F1B stimulus from responses to the 2F1B stimulus again isolated neural responses to the second flash, but with a multi-modal context (when it was immediately preceded by a combined flash/beep). Further contrasting these difference waves thus indicated differences in the response to the second flash, depending on whether it was preceded by a uni- or multi-modal context stimulus. This potentially ‘lasting’ effect of multi- vs uni-modal stimuli on visual processing could explain the presence of the illusion.

6.2 Methods

6.2.1 Participants

Fourteen participants were recruited from the student body at the Brain Sciences Institute at Swinburne University. Ages ranged from 22-32 years (M = 28.3, SD = 3.4), and four were male. Six had participated in the previous experiment reported in Chapter 5, although the testing session was approximately one year later. The protocol was approved by the Swinburne University Human Experimentation Ethics Committee (see Appendix 2), and written, informed consent was obtained from all participants. Data from two participants was discarded due to equipment error related to the recording of stimulus triggers, and from one participant due to a lack of trials for one response type (see Section 6.4.2), leaving a total of 11 participants.
6.2.2 Stimuli and apparatus

The stimuli and apparatus were the same as those used in Chapter 5, except that the refresh rate of the monitor was increased to 75 Hz, and the timing of the auditory and visual stimuli were slightly adjusted such that they occurred simultaneously, rather than with the 23ms delay used in Chapter 6. This stimulus timing is simpler to implement and interpret, especially in the context of ERP analysis, and has since been used successfully in fMRI (Watkins et al., 2006) and ERP (Meylan and Murray, 2007) investigations of the flash-beep illusion.

The experiment was conducted in a quiet, sound-treated, and electrically shielded room located at the Brain Sciences Institute. The background sound level was approximately 39 dB SPL (A-weighted). The visual stimulus consisted of a white disk, which flashed once or twice on a 15-inch cathode ray tube monitor (CRT) with a black background. The disk subtended 3° of visual angle and was located 7.5° below a fixation cross, which was positioned 2.5° above the centre of the screen. Each flash consisted of two refresh periods (26.67 ms, see Figure 1). On selected trials, short beeps were presented, simultaneously with the flashes, from a speaker placed centrally under the CRT. The beep was a 3500 Hz, 83 dB SPL (A-weighted) sine wave of 8 ms duration, with 3 ms rise and fall times. Participants sat in a comfortable chair with the head 100 cm from the CRT display. A keyboard rested on a desk at a comfortable distance directly in front of the CRT. The auditory and visual stimuli were controlled using Presentation 10.1 (Neurobehavioural Systems) running on a standard Windows XP PC. Auditory stimuli were generated using a SoundBlaster Audigy 2 ZS 24-bit PCI soundcard (Creative Resource, Singapore) in “Direct X Hardware” mixing mode. Using a cathode-ray oscilloscope, light-pen and microphone, the average delay between the visual flash stimulus and the first measurable peak of the auditory stimulus was measured 20 times. The mean delay of the auditory stimulus was .3 ms, with no measureable standard deviation (all measurements were .3 ms, despite efforts to disrupt the timing by launching other software programs, conducting file searches etc).
6.2.3 Electrophysiological recordings

The continuous EEG was recorded from 60 sintered silver/silver-chloride electrodes mounted in an elastic cap (Compumedics Quick-Cap) according to the international 10-20 system. The electrodes were FP1, FPZ, FP2, AF3, AF4, F7, F5, F3, F1, FZ, F2, F4, F6, F8, T7, FC5, FC3, FC1, FC2, FC4, FC6, FT8, T8, C5, C3, C1, CZ, C2, C4, C6, T8, CP7, CP5, CP3, CP1, CPZ, CP2, CP4, CP6, TP8, P7, P5, P3, P1, PZ, P2, P4, P6, P8, PO7, PO5, PO3, POZ, PO4, PO6, PO8, O1, OZ, O2. The continuous EEG was amplified, filtered with a bandpass of 0.1-100 Hz, and digitised at a sampling rate of 1000 Hz using a Synamps II EEG amplifier (Compumedics Neuroscan, Melbourne). Event codes were sent to the EEG amplifier from the parallel port of the PC running the stimulus presentation software using a custom-made cable. All stimulus and response events were assigned a code and recorded along with the continuous EEG. To enable checking of the stimulus timing, all auditory and visual sub-components of each stimulus were assigned a separate code. For example, a 1F1B stimulus resulted in the recording of two separate event codes. Analysis of the event code timing revealed no inconsistencies. The continuous EEG was saved in Neuroscan ‘cnt’ format.

6.2.4 Procedure

The procedure was the same as that outlined in Chapter 5, except that a reduced number of flash/beep stimulus combinations was used in order to reduce the length of the experiment, while increasing the number of repetitions of each trial type sufficiently for EEG analysis. The multi-modal fission and fusion illusion stimuli were presented, along with uni-modal control stimuli (a single or double flash), and congruent multi-modal stimuli, where the number of flashes and beeps was equal. In each trial, there was either a single or double flash, along with zero, one, or two beeps. Trials types will be henceforth referred to using a code indicating the number of flashes followed by the number of beeps – “2F2B” thus refers to a trial with two flashes and two beeps. The six possible trials types were therefore 1F0B, 1F1B, 1F2B (fission illusion), 2F0B, 2F1B (fusion illusion) and 2F2B. In trials with multiple flashes or beeps, the time between the onsets of
successive beeps or flashes was 66.7 ms, corresponding to 5 CRT refresh periods. An example of a 2F2B trial is shown in Figure 7-1.

Figure 6-1. Flash beep stimulus timing. Shown is an example of a 2-flash, 2-beep (2F2B) trial.

The fixation cross was displayed alone for an interval that varied randomly in each trial between 1200 and 1500 ms. This random variation was introduced in order to reduce the possibility of participants predicting the stimulus onset and responding too quickly, and to reduce the possibility of readiness or contingent negative variation potentials occurring in the pre-stimulus EEG (Walter et al., 1964, Teder-Salejarvi et al., 2002). The flash/beep sequence then began. Following the sequence was another short randomly varied interval (1200 to 1500 ms), after which the text “How many flashes did you see?” was displayed in place of the fixation cross. This text remained in place until the participant made a response on the keyboard, or until 2.5 seconds passed, after which time the response was deemed invalid. Participants were instructed to keep their gaze on the fixation cross during each trial and count the number of flashes that would appear whilst ignoring the beeping sounds. The response was made after each trial by pressing keys labelled ‘1’ or ‘2’ on a keyboard.

The six possible flash/beep stimuli were presented in random order 20 times each in a single block. This block was repeated 5 times (with trials re-randomised each time). Each stimulus was thus presented 100 times. Keyboard responses were recorded for each stimulus presentation. Each block ran for an average duration of 10 minutes and breaks could be taken between each block. The total testing time was approximately 1.5 hours, including breaks, instruction time, and EEG setup and cleanup time.
6.3 Analysis and results – behavioural data

6.3.1 Flash-beep task – mean responses.

Although accuracy scores will be analysed in the remainder of this chapter, the mean responses are shown in Figure 7-2 to allow easy comparison with the mean response graphs in the previous chapter. In single-flash trials, the mean number of flashes reported was close to 1 when there were 0 or 1 accompanying beeps, indicating that participants had no difficulty perceiving the single flash stimulus with a congruent or absent beep. However, when 2 beeps were present, the reported number of flashes increased dramatically (fission illusion). In two-flash trials, the number of reported flashes was close to 2 when either 0 or 2 beeps were presented, indicating that participants also had no difficulty perceiving the double-flash stimulus. When a single beep was presented along with the double flash, however, the mean response fell, although the difference was smaller than in the one-flash trials. No statistical analysis was carried out on the mean response data, as accuracy rates will be analysed.

Figure 6-2. Averaged responses (± SEM) for each stimulus type in one-flash trials (left) and two-flash trials (right).

6.3.2 Flash-beep task – accuracy

Accuracy scores were analysed in order to firstly determine whether participants could accurately count the visual flash stimuli either in the absence of
an auditory stimulus or with a congruent number of beeps, and secondly to
determine the extent of fission and fusion illusions reported in trials where an
illusion was expected.

For each of the six stimulus types, accuracy scores were calculated by
dividing the number of correct responses by the total number of responses made,
so that non-responses (when the participant took longer than 2.5 seconds to
respond) were not counted as incorrect. Non-responses were rare (99.1% valid
responses). For the non-illusion trials (1F0B, 1F1B, 2F0B, 2F2B), the accuracy
scores reflected the degree to which participants were able to accurately count the
visual flash stimuli with no beeps or with a congruent number of beeps.
Conversely, for the illusion trials, low accuracy scores indicated the presence of
illusory perception. In fission trials (1F2B), low accuracy indicated that more flashes
were reported than were presented, and in fusion trials (2F1B), low accuracy
indicated that less flashes were reported than were presented.

Figure 6-3 shows mean accuracy scores for each stimulus type. In general,
all participants responded with a high level of accuracy for all non-illusion stimuli,
suggesting that the visual stimuli were not ambiguous and that the visual flashes
could be counted relatively easily. This was the case both in the uni-modal trials
with no auditory stimulus present (1F0B and 2F0B) and in multi-modal congruent
trials where the number of auditory and visual stimuli were equal (1F1B and
2F2B). However, accuracy was lower for both types of illusion trials (fission –
1F2B, and fusion – 2F1B).
Figure 6-3. Mean accuracy measures (± SEM) for each stimulus type. In fission illusion trials (1F2B), participants often reported more flashes than were present; in fusion illusion trials (2F1B), participants often reported less flashes than were presented.

The significance of these effects was assessed using a repeated measures analysis of variance (ANOVA) with within-subjects factors for Nflash (1 flash, 2 flashes), and Nbeep (0 beeps, 1 beep, 2 beeps). There were significant main effects of Nflash, *F*(1,11) = 10.9, *p* = .007, \( \eta^2 = .50 \), and Nbeep, *F*(2,22) = 8.8, *p* < .002, \( \eta^2 = .44 \), as well as a significant Nflash x Nbeep interaction, *F*(2,22) = 18.9, *p* < .00, \( \eta^2 = .63 \). Accuracy in the illusion trials compared to the corresponding non-illusion trials was examined by decomposing the interaction using simple effects analysis (Howell, 2009). Pairwise comparisons with Sidak adjusted alpha levels are reported throughout. In one-flash trials, accuracy scores were lower in fission illusion trials compared with both the uni-modal (1F0B), *p* = .003, and congruent multi-modal
control trials (1F1B), \( p = .002 \). In two-flash trials, accuracy in the illusion trials (2F1B) was lower than in the congruent multi-modal trials (2F2B), \( p = .02 \), but not compared to the uni-modal trials (2F0B), \( p = .06 \).

As the number of males (\( N = 4 \)) and females (\( N = 8 \)) were unequal, 2-sample t-tests were performed on accuracy scores for each stimulus type, in order to test for any bias in accuracy based on gender. Accuracy scores were not significantly different between males and females for any stimulus.

Thus, both the flash-beep fission and fusion illusions were present, although the fission illusion was significantly stronger, and performance in the 2F1B condition (where the fusion illusion might be expected to occur) was still over 75% on average.

### 6.4 EEG - Methods of analysis

#### 6.4.1 EEG pre-processing procedures

A main advantage of the flash-beep illusion paradigm is the possibility of comparing brain responses to the same physical stimulus depending on whether or not the illusion was reported. However, the number of illusion trials available to construct the average ERP in each case is necessarily divided by the proportion of responses in each case. A participant who reported an illusion on 80% of fission trials, for example, would have 80% of the total number of fission trials available to calculate the ‘illusion’ response, and only 20% for the non-illusion response. As there were a maximum of 100 trials available for each stimulus type, this would result in 80 trials available for the ‘illusion’ response and only 20 for the ‘no-illusion’ response. These 20 trials are unlikely to be sufficient to construct an average ERP, especially as some extra proportion of those trials would likely be rejected during pre-processing steps. It was thus critical to retain as many trials as possible in the EEG analysis. The EEG pre-processing procedures thus emphasised the correction of EEG artefacts such as blinks using ICA techniques, with the minimum of trials rejected.
A MATLAB script was developed for pre-processing the continuous EEG. It contained a mixture of basic MATLAB functions, as well as functions from the EEGLAB (Delorme and Makeig, 2004) and ADJUST (Mognon et al., 2010) toolboxes. The continuous EEG was imported from Neuroscan ‘cnt’ format into MATLAB, and saved in EEGLAB ‘set’ format. While the EEG was still stored at a high sampling rate (1000 Hz), the auditory and visual stimulus event marker timings were checked for synchronisation and consistency. Auditory and visual stimulus markers were found to occur at their expected times within the time resolution of the EEG sampling (1 ms per sample). The continuous EEG was then filtered using a second-order bi-directional (zero-phase-shift) Butterworth bandpass filter with cutoff frequencies of 0.1-45 Hz and slopes of 12dB/oct. Following filtering, the EEG was re-sampled at 500 Hz.

The stimulus and response event codes were then simplified to allow for easier analysis – the first stimulus in a group was re-coded according to the overall stimulus type, and all other stimulus codes were deleted, leaving a single event code per stimulus, with six possible event types (1F0B, 1F1B, 1F2B, 2F0B, 2F1B, 2F2B). The stimulus-response contingencies are detailed in Table 6-1.
Table 6-1. Stimulus-response contingencies. For non-illusion stimuli (1F0B, 1F1B, 2F0B, 2F2B), incorrect responses were not analysed. For illusion stimuli (1F2B, 2F1B), separate codes were assigned depending on the response.

<table>
<thead>
<tr>
<th>Stimulus</th>
<th>Response</th>
<th>Code</th>
</tr>
</thead>
<tbody>
<tr>
<td>1F0B</td>
<td>1</td>
<td>1F0B</td>
</tr>
<tr>
<td>1F0B</td>
<td>2</td>
<td>-</td>
</tr>
<tr>
<td>1F1B</td>
<td>1</td>
<td>1F1B</td>
</tr>
<tr>
<td>1F1B</td>
<td>2</td>
<td>-</td>
</tr>
<tr>
<td>1F2B</td>
<td>1</td>
<td>1F2B-no_ill</td>
</tr>
<tr>
<td>1F2B</td>
<td>2</td>
<td>1F2B-ill</td>
</tr>
<tr>
<td>2F0B</td>
<td>1</td>
<td>-</td>
</tr>
<tr>
<td>2F0B</td>
<td>2</td>
<td>2F0B</td>
</tr>
<tr>
<td>2F1B</td>
<td>1</td>
<td>2F1B-ill</td>
</tr>
<tr>
<td>2F1B</td>
<td>2</td>
<td>2F1B-no_ill</td>
</tr>
<tr>
<td>2F2B</td>
<td>1</td>
<td>-</td>
</tr>
<tr>
<td>2F2B</td>
<td>2</td>
<td>2F2B</td>
</tr>
</tbody>
</table>

Once segmented, artefact rejection and correction procedures were applied. A three-step approach was taken:

1) EEG epochs and channels containing artefacts were rejected. Only epochs which were considered to be severe, or of a unique ‘stereotyped’ nature likely to cause an unsuccessful ICA decomposition (see Chapter 4) were rejected.

2) ICA was run, and components representing ocular or other artefacts were marked and removed.

3) The data was re-projected back into EEG epochs, and previously deleted channels re-interpolated.

Statistical thresholding was used to identify EEG electrodes with artefact contamination. Channels that had frequency spectrums in two separate ranges that were outside statistical thresholds were identified. The two ranges were 0-5 Hz, used to identify ‘flatline’ channels as well as those with slow drifts caused by sweat or other battery potential type artefacts, and 35-100 Hz, to identify channels with muscle movement or other high-frequency noise. Channels with frequency spectrums more than ± 2 standard deviations (for the low frequency range) or 3 standard deviations (for the high frequency range) from the mean of all channel
spectrums were marked as bad. The identified channels were checked visually before being finally rejected. Figure 6-5 shows histograms of the mean power found in the electrodes in the 0-5 Hz range (top left) and 35-100 Hz range (top right). The rejection thresholds are shown by the red lines. In this participant, channels 6 and 28 were rejected as they had excessive high low-frequency power, as can be seen in the three example epochs in the lower panel. The rejected channels are shown in red. Between 2 and 6 channels ($M = 4.14, SD = 1.35$) were rejected for each participant. The rejected channels were noted and removed from the EEG.

4 Although the continuous EEG was bandpass filtered from 0.1-45 Hz in a previous step, the slope of this filter was only 12 dB/octave. The 35-100 Hz range used in this processing step overlaps with the previous filter, but enough high-frequency energy remained firstly to cause artefacts, and secondly to allow detection of abnormal levels of activity.
Figure 6-4. Example of statistical thresholding technique used to determine bad electrodes. In this example, electrodes 6 and 28 were marked for rejection, as they had mean power outside ± 2 standard deviations from the mean of all electrodes.

EEG epochs containing artefactual contamination were also rejected using statistical thresholding. As a first step, epochs with voltages ±500 μV were marked as bad. An iterative procedure implemented in the EEGLAB ‘pop_autorej’ function then identified epochs that contained data points more than 6 standard deviations from the mean. If less than 3% of the total number of epochs were marked for rejection, the marked epochs were rejected, and the procedure repeated without the marked epochs contributing to the calculations. If the number of marked epochs was more than 3% of the total number of epochs, no epochs were rejected, the standard deviation threshold was increased by .5 and the procedure repeated. This process continued until no more epochs contained data values outside the threshold, or until 8 iterations were completed. After the final
set of epochs was marked for rejection, all epochs were inspected visually and
minor adjustments made before finally rejecting. From 26 to 118 epochs \( (M = 62.3, SD = 27.1) \) out of the total 600 epochs were rejected for each participant using this technique.

Once a set of epochs free of stereotyped artefacts was established for each participant, the data was re-referenced to the average of all channels, and extended infomax ICA (Bell and Sejnowski, 1995, as implemented in EEGLAB, Delorme and Makeig, 2004) was run on the concatenated 1 s epochs. A combination of statistical thresholding and visual inspection was used to identify components related to eye-blinks, eye-movements, and other artefacts. The thresholding procedure was carried out using the ‘ADJUST’ plugin for EEGLAB (Mognon et al., 2010). This procedure identified components with stereotyped temporal, spatial and joint spatio-temporal features (see Section 4.7 in Chapter 4 for a review of the use of ICA in correcting EEG artefacts). Components with features that exceeded a threshold computed by examining the whole set of components from each participant were then marked for rejection by the plugin. The time course, spectrum, event-related average, and spatial map for all components, including those suggested for rejection by the ADJUST plugin, were visually examined. Those related to eye-blinks and eye-movements, high-impedance electrodes or other gross artefacts were removed.

After the artefact-related components were removed, the data was firstly back-projected into EEG channel time-series data, and the EEG channels that were previously removed were interpolated from neighboring electrodes using a spherical spline model (Perrin et al., 1987) implemented in the ‘eeg_interp.m’ EEGLAB function (Delorme and Makeig, 2004). Secondly, EEG epochs for each stimulus/response combination were saved into separate EEGLAB ‘set’ files. Finally, the average of all trials (the ERP) for each electrode, and the standard deviation across electrodes (equivalent to the mean global field power – GFP, see Chapter 4) was calculated and stored on disk as well as in a separate MATLAB cell array for each participant and stimulus-response combination.
6.4.2 Calculation of UNI and MULTI difference waves

The analysis of the EEG data collected during this paradigm posed special problems. It was initially anticipated that ERP responses could be compared between ‘illusion’ and ‘no-illusion’ responses to the two illusion stimuli – that is, responses to the same stimuli (1F2B and 2F1B), depending on whether the participant indicated an illusory perception or not. However, despite a data pre-processing approach that emphasised the correction or interpolation of data rather than outright rejection of trials that were contaminated with eye-blinks or muscle activity etc, there were insufficient trials to enable this type of analysis for most participants. For fission (1F2B) stimuli, there were four participants who had sufficient remaining trials in ‘illusion’ and ‘no-illusion’ responses. Table 6-2 shows the number of ‘good’ EEG epochs remaining for analysis after all pre-processing stages were complete. As can be seen, for both the 1F2B and 2F1B stimuli, many of the participants had only a very small number of trials available in some trial-response categories.

<table>
<thead>
<tr>
<th>ID</th>
<th>1F0B</th>
<th>1F1B</th>
<th>1F2B-no ill</th>
<th>1F2B-ill</th>
<th>2F0B</th>
<th>2F1B-no ill</th>
<th>2F1B-ill</th>
<th>2F2B</th>
</tr>
</thead>
<tbody>
<tr>
<td>s02</td>
<td>96</td>
<td>88</td>
<td>91</td>
<td>5</td>
<td>96</td>
<td>85</td>
<td>3</td>
<td>90</td>
</tr>
<tr>
<td>s04</td>
<td>96</td>
<td>84</td>
<td>85</td>
<td>2</td>
<td>86</td>
<td>89</td>
<td>3</td>
<td>90</td>
</tr>
<tr>
<td>s05</td>
<td>92</td>
<td>94</td>
<td>91</td>
<td>6</td>
<td>95</td>
<td>91</td>
<td>0</td>
<td>92</td>
</tr>
<tr>
<td>s08</td>
<td>79</td>
<td>81</td>
<td>81</td>
<td>80</td>
<td>89</td>
<td>75</td>
<td>3</td>
<td>88</td>
</tr>
<tr>
<td>s09</td>
<td>83</td>
<td>97</td>
<td>44</td>
<td>42</td>
<td>82</td>
<td>27</td>
<td>68</td>
<td>88</td>
</tr>
<tr>
<td>s10</td>
<td>68</td>
<td>84</td>
<td>45</td>
<td>39</td>
<td>75</td>
<td>81</td>
<td>3</td>
<td>67</td>
</tr>
<tr>
<td>s11</td>
<td>81</td>
<td>74</td>
<td>37</td>
<td>47</td>
<td>87</td>
<td>82</td>
<td>1</td>
<td>85</td>
</tr>
<tr>
<td>s12</td>
<td>87</td>
<td>91</td>
<td>82</td>
<td>9</td>
<td>89</td>
<td>94</td>
<td>0</td>
<td>91</td>
</tr>
<tr>
<td>s13</td>
<td>91</td>
<td>89</td>
<td>28</td>
<td>69</td>
<td>95</td>
<td>73</td>
<td>24</td>
<td>94</td>
</tr>
<tr>
<td>s14</td>
<td>88</td>
<td>87</td>
<td>14</td>
<td>76</td>
<td>85</td>
<td>89</td>
<td>1</td>
<td>83</td>
</tr>
<tr>
<td>s15</td>
<td>75</td>
<td>89</td>
<td>2</td>
<td>90</td>
<td>87</td>
<td>58</td>
<td>28</td>
<td>87</td>
</tr>
</tbody>
</table>

Table 6-2. Number of trials remaining for each participant in each stimulus/response combination. Most participants had insufficient trials in either ‘illusion’ or ‘no-illusion’ responses to fission (1F2B) stimuli to allow a direct comparison of illusory vs non-illusory responses to the same stimulus.

As alluded to in the introduction, an alternative analysis was thus performed, partly following the methods described in Meylan and Murray (2007),
who faced a similar problem. In this approach, only four stimulus types were analysed – uni-modal stimuli, which were the 1F0B and 2F0B stimuli, and multi-modal stimuli, which were the 1F1B and 2F1B stimuli. As will be described in more detail later, the response to the second flash only was isolated by defining UNI and MULTI difference waves as follows: UNI = 2F0B – 1F0B, and MULTI = 2F1B – 1F1B. Figure 6-5 shows how the UNI and MULTI difference waves were calculated. The response to the second flash was thus isolated, depending on whether it was immediately preceded by a uni-modal flash stimulus, or a multi-modal flash/beep stimulus. In this way, the effect of either a uni-modal or multi-modal preceding context on the response to the second flash could be determined.

![Figure 6-5. Diagram showing the derivation of UNI and MULTI difference waves from the uni-modal (1F0B and 2F0B) and MULTI (2F1B and 1F1B) stimulus types.](image)

Both the UNI and MULTI difference waves represent the neural response to only the second flash. The analysis of the UNI and MULTI difference waves was therefore designed to determine the possible timing and location of statistically significant differences between the two waves. Rather than focus on the identification and measurement of ERP components (such as the N1, P2, etc), the analysis sought firstly to determine time points at which the statistically significant differences occurred, and secondly to use source localisation methods.
to locate the parts of the cortex in which these differences occurred. As explained in Chapter 4, the global field power (GFP) is a measure of the overall electrical field response at the scalp for each time point. It has the advantage that it is not affected by the choice of reference electrode, and the corresponding disadvantage that it provides no information about the possible neural sources underlying the electric field measured at the scalp. Hence in the current study, it was used to determine time periods at which the overall electrical response to the UNI and MULTI stimuli differed from each other, without requiring any prior assumptions of electrode location or reference configuration. The sources underlying the electric response in these time periods were then estimated using source localisation techniques.

The two-step process is outlined below:

1) *When* were there significant differences between the UNI and MULTI differences waves?
   a. **Visualisation** - grand average waveforms calculated for the 1F0B, 1F1B, 2F1B, and 2F2B stimuli.
   b. **Visualisation** - calculation and point-wise significance testing of the UNI and MULTI difference waves.
   c. **Analysis** of significantly different time intervals - calculation and point-wise significance testing of the GFP.

2) *Where* in the brain could those differences have been generated?
   a. **Visualisation** - sLORETA localisation of current source density for UNI and MULTI waves.
   b. **Statistical analysis** of differences in sources during the time intervals previously found.
6.4.3 When - determining time periods when statistically significant differences occurred

Although the final determination of time intervals was calculated using the GFP, the scalp ERP data for the four stimuli, as well as the UNI and MULTI differences waves, was also visualized to allow comparison with previous work. After the average ERP and GFP were calculated separately for each stimulus type, the statistical significance of differences between the ERPs in each condition were assessed by using point-wise non-parametric multiple permutation tests, with 2000 permutations, for each time point and electrode. Throughout, similar permutation tests were used to test for statistically significant differences between pairs of waveforms. This method allowed the visual identification of periods of statistically significant differences between waveforms in a manner more conservative than standard parametric t-tests, as no assumptions of normality were required. In addition, only periods of significant differences longer than 10 samples (20 ms) were considered reliable (Guthrie and Buchwald, 1991). An intensity plot of the critical \( t \) values (two-tailed, for \( df = 10 \)) on a surface defined by the time points and electrodes provided a visual impression of the time periods and scalp electrodes where significant differences occurred. However, it should be emphasized that the both the ERP voltage plots and \( t \)-statistic intensity plot only provided information regarding the time periods and scalp electrodes at which significant differences were found using the particular reference configuration in this study (the average reference), and are not generalizable. For reasons of display consistency, the x-axis on all time-based plots (ERPs and GFP) will represent time relative to the first flash. In the case of the UNI and MULTI difference waves, in which the response to the first flash was subtracted away, the reader will need to take this time into account. The second flash was always presented at 67 ms, and its location will be marked on the plots. In some cases, times relative to both the first and second flash will be provided.

In order to determine appropriate time periods in which to localize the cortical sources of the differences, the reference-free GFP measure was used. The
GFP was determined for each participant by calculating the standard deviation across all electrodes at each time point. This is equivalent to the sum of squares of potential differences at all possible electrode combination, and indicates the overall strength of the electric field at the scalp at each time point. To calculate the grand average GFP, the mean of all participant’s GFP waveforms was calculated. To determine the time intervals at which the GFP for UNI and MULTI waves was statistically significantly different, a point-wise permutation test with 2000 permutations and a $p$ value of 0.05 was used. In addition, only periods of significant differences longer than 10 samples (20 ms) were considered reliable (Guthrie and Buchwald, 1991).

### 6.4.4 Where - source analysis of statistically significantly different time periods

In order to determine the probable location of cortical generators underlying the differences in ERP field strength measured at the scalp, Standardised Low-Resolution Electromagnetic Tomography (sLORETA, Pascual-Marqui, 2002, Pascual-Marqui et al., 1994, see Chapter 4 for review of technique) was used to estimate the cortically-constrained current source density of the UNI and MULTI difference waves for each participant. The standardised current density was calculated at 6239 voxels (Lancaster et al., 2000) at 5mm resolution, using a realistic lead field model (Fuchs et al., 2002). Standard electrode locations from Oostenveld and Praamstra (2001) were used. The sLORETA software (Key Institute) was used for the source estimation as well as the statistical testing of the source differences between UNI and MULTI waves (Pascual-Marqui, 2002).

sLORETA solutions for all time points and all participants were first calculated for the UNI and MULTI waves. For a given time segment that was revealed as significant by the GFP permutation test, average sLORETA images were calculated for the UNI and MULTI waves for each participant by taking the average of all instantaneous sLORETA images within the significant time segment. To determine the statistical significance of differences between localisations for the UNI and MULTI difference waves, ‘Statistical non-Parametric Mapping’ (SnPM)
was used, as implemented in the sLORETA software (Pascual-Marqui, 2002). SnPM performed voxel-wise randomisation tests (5000 permutations), and calculated critical thresholds and \( p \)-values corrected for the number of multiple comparisons involved in the voxel-wise test. The log ratio of averages (similar to the \( F \)-statistic) was calculated for every voxel, and thresholded with alpha level of 0.05. For full details of the SnPM method, including reliability analysis and assessments of the performance of corrections for multiple comparisons, see Holmes et al (1996) and Nichols et al (2002). The end result of the SnPM method, as implemented in the sLORETA software, was a thresholded map of the probability of the source of differences between the UNI and MULTI difference waveforms. The MNI co-ordinates of the location of the maximum or minimum pseudo-\( F \) statistic was converted to a brain region using the Talairach map (Lancaster et al., 2000).

### 6.5 EEG Results – multisensory context analysis

#### 6.5.1 Waveform analysis - grand averages of uni- and multi-modal stimuli

The analysis focussed first on determining differences in the timing, magnitude, and topography differences between ERPs in response to the uni-modal and multi-modal stimuli. Figure 6-5 shows the grand average ERPs at all electrode sites, calculated for the uni-modal (1F0B and 2F0B) and multi-modal (1F1B and 2F1B) stimuli. The periods of time where the two waveforms significantly differed across the group of participants are highlighted with grey boxes.
Figure 6-6. Grand average ERPs for UNI (top panel) and MULTI (bottom panel) stimuli. Significant differences between the waveforms are marked.
with grey boxes (permutation test, $p < .05$). The x-axis represents time relative to the first flash.

Visual inspection, combined with the exploratory permutation tests, revealed that the grand average waveforms for the uni-modal stimuli (1F0B and 2F0B) show the expected visual evoked potential (VEP) morphology, with strong P1 and N1 peaks visible, especially at the occipital electrodes (Fig 6-6, top panel). Note that this data is displayed with the average reference, rather than the nose reference often used in VEP studies. Also visible are differences between the 1F0B (red) and 2F0B (green) ERPs. Significant differences between responses to the two stimulus types are indicated with the grey boxes. Significant differences, likely reflecting the response to the second flash at 66.7 ms, occurred in central, parieto-occipital and fronto-central sites from around 220-300 ms post-stimulus.

ERPs in response to the multi-modal stimuli (1F1B and 2F1B) show a similar overall morphology to the uni-modal stimuli, although overall amplitudes are larger, probably reflecting the summation of auditory and visual ERPs. There were again differences between the waveforms at similar latencies to the uni-modal stimuli, as well as at earlier latencies (note the positive peak at around 150 ms). The spatial pattern of significantly different intervals appeared more broad than with the UNI stimuli, with more electrodes showing significantly different time intervals. The intervals were also more broadly spread in time, extending both earlier and later.

From statistically-guided visual inspection, it thus appeared that the response to the second flash may have been larger when it was preceded by a multi-modal stimulus compared to a uni-modal stimulus.

### 6.5.2 Statistical comparison of ‘UNI’ and ‘MULTI’ difference waves

In order to statistically compare responses to the second flash as a function of the preceding uni-modal or multi-modal stimulus, the UNI and MULTI
difference waveforms were calculated. A multi-step approach was taken (see Analysis section 6.4.2) to determine the timing and neural sources of the differences between responses to the second flash when preceded by either a uni-modal (UNI difference wave) or multi-modal (MULTI difference wave) stimulus.
Figure 6-7. Grand averages for UNI (red) and MULTI (green) difference waves. Grey boxes show time points at which a permutation test indicated a significant difference between the waveforms. The x-axis represents time relative to the first flash.

Figure 6-7 shows the UNI and MULTI difference waveforms averaged across all participants. Recall that both waveforms represent the isolated response to the second flash, with the only difference being the immediately prior stimulus. The UNI waveform shows the response to the second flash when it was preceded by another flash stimulus, while the MULTI waveform shows the response to the second flash when it was preceded by a multi-modal flash/beep stimulus. Any differences between these two waves were thus due to the differential effect of the preceding stimulus. There were several electrodes where such differences were evident. At fronto-central sites, the MULTI waveform was larger than the UNI waveform from 0-100 ms, whereas at occipital and parietal sites, the opposite was the case, and additional later differences were also present. There was a reasonably
complex pattern of early and late differences across a variety of mostly central, parietal, and occipital electrodes.

**6.5.3 Determining time periods of significantly different electrical responses between UNI and MULTI waves**

In order to simplify the display of the UNI and MULTI ERP data, and display consistent time intervals where the overall electric field response differed between the two waves, the \( t \)-statistic from a point-wise \( t \)-test between the UNI and MULTI waves for all electrodes was plotted on a common axis, and the GFP was analysed. Figure 6-8 (top panel) shows the results. Only \( t \)-values larger than the critical \( t \)-value for 10 degrees of freedom and extending over at least 10 consecutive time points (20 ms) are shown. This display is essentially a repeat of the data shown in Figure 6-6, but all electrodes are shown on a common x-axis, and critical \( t \)-values testing the difference between the waveforms are shown rather than separate voltages for each waveform. As can be seen, there were three main intervals over which differences occurred.

The first was a very early difference from 0-40 ms at central and lateral sites. These differences were before the presentation of the second flash (at 67 ms) and were not analysed further. From 120-250 ms there followed a large number of differences across many electrodes. At frontal sites the UNI waveform was mostly larger than the MULTI waveform (red clusters), whilst the opposite was the case at more parietal electrodes (blue clusters). There were also later differences, particularly a cluster of occipital sites where the MULTI response was larger than the UNI response. As has been reviewed previously (Chapter 4), determining significant time intervals directly from electrode voltage data is problematic, as the responses are dependent on the choice of reference electrode.

The grand average GFP for the UNI (red line) and MULTI (green line) is shown in the middle panel of Figure 6-8. The bottom panel shows the \( p \)-values from a point-wise permutation test between the two GFP waveforms. There are
two features of note. Firstly, it is clear that overall, the MULTI difference wave showed a larger response strength than the UNI difference wave at all timepoints. Secondly, the differences were significant in two main intervals – an early interval from 120-190 ms, and a later interval from 300-320 ms. Thus, the overall electric field measured at the scalp in response to the second flash was significantly larger in two specific intervals when it was preceded by a multi-modal flash/beep stimulus compared to when it was preceded by a flash alone.
Figure 6-8. Results from statistical testing of UNI vs MULTI difference waves. The x-axis on all plots represents time relative to the first flash. The grey line at 67 ms indicates the time of the second flash. TOP: P-values from point-wise permutation testing between the UNI and MULTI difference waves at every electrode. Electrodes are arranged in bands separated by grey horizontal lines from the frontal (top of plot) to occipital (bottom of plot) regions. Within each band, electrodes are arranged from left-most (top of band) to right-most (bottom of band). The midline electrode in each band is shown on the y-axis. For example, in the lowest (occipital) band, the three electrodes shown are O1, Oz and O2. MIDDLE: The mean (across participants) global field power (MGFP) calculated across all electrodes for UNI and MULTI difference waves. The MULTI waves shows increased activity in early and late time windows. BOTTOM: p-values from a point-wise permutation test between the two GFP waveforms show that the MULTI difference waveform is significantly larger for a very short period at around 50 ms, and extended periods from 100-200 ms, and from 300-350 ms.

Before performing the source localisation analysis, topographic maps were made from the grand average UNI and MULTI waves. The mean voltages at each electrode in the early and late time intervals were calculated and plotted on a 2-D head map using interpolation over a fine Cartesian grid. Note that the early interval was reduced from 120-190 ms to 130-160 ms in order to make the durations of the early and late intervals more similar. The 30 ms interval chosen was in the centre of the early period. Figure 6-9 shows topographic maps for grand average UNI and MULTI grand average waves in the early and late time intervals. T-tests (2-tailed, $df = 10$) were also performed at each electrode comparing the mean voltage over the specified interval at each electrode. The topographic maps and t-tests were consistent with the results from the t-intensity plot, and showed that differences in the early interval were present at the parietal and lateral electrodes, and that the later differences were at the parietal and occipital electrodes.
Figure 6-9. Topographic maps of the UNI and MULTI difference waves in the early (top row) and late (bottom row) time intervals. Electrodes at which the mean voltage in each range was statistically significantly different in each condition are indicated with a red dot ($p < .05$, $df = 10$).

6.5.4 Localisation of UNI vs MULTI differences

The analysis of the ERP and GFP waveforms showed significant differences between the UNI and MULTI waves in two distinct time intervals – an early phase from 130-160 ms, and a late phase from 300-320 ms. Exploratory analysis via $t$-intensity plots and topographic maps calculated over the two intervals showed that differences between the UNI and MULTI waveforms were found at central and parietal electrodes during the early interval, and at parietal and occipital electrodes during the late interval. In order to more directly specify the brain regions involved in these differences, source localisation using sLORETA was performed on the UNI and MULTI difference waves. sLORETA current source density (CSD) maps were calculated for the UNI and MULTI waves for each participant. These were then averaged across the early and late time intervals, and
SnPM (see Chapter 4) statistics calculated to determine any statistically significant differences in the sources underlying the UNI and MULTI waves.

Figure 6-10 shows the sLORETA CSD maps plotted on the MNI template brain for the UNI and MULTI responses in the late interval (only the late interval was displayed for brevity), as well as the difference between the two (MULTI - UNI). As expected from the previous analyses, the main sources for both were in the parietal and occipital lobes. Also as expected, the CSD values for the MULTI responses were generally higher than for UNI responses. Subtracting the UNI from the MULTI CSD maps (Figure 6-10, bottom row) shows mainly positive differences, indicating that the MULTI maps had higher CSD values. The differences were mainly focussed in the occipital lobes.
Figure 6-10. sLORETA source localisation results (µA/mm³) displayed on the MNI T2 template brain for the MULTI (top row) and UNI (bottom row) waves, in the late interval (300-320 ms). The difference between the two (MULTI-UNI) is shown in the bottom row. Four views are shown: from left to right these are top, back, left, and right views.

In order to determine the statistical significance of the source differences between the MULTI and UNI waves, the sLORETA images for each participant, averaged across each timeframe, were compared between the MULTI and UNI responses using the sLORETA built-in voxel-wise randomisation tests based on SnPM techniques (see Analysis section and Chapter 4). The pseudo F-statistics for voxels with significant differences were plotted in their appropriate locations in Talairach space on the MNI ‘Colin27’ T2 template brain (Holmes et al., 1998). Figures 6-11 and 6-12 show these statistical difference maps for the early and late time intervals respectively.
Figure 6-11. sLORETA statistical image showing significant pseudo-$F$ values from a voxel-wise random permutation test between MULTI and UNI sLORETA source localisations in the early timeframe (130-160 ms, or 63-93 ms after the second flash).
Figure 6-12. sLORETA statistical image showing significant pseudo-\( F \) values from a voxel-wise random permutation test between MULTI and UNI sLORETA source localisations in the late timeframe (300-320 ms, or 233-253 ms after the second flash).

In the early timeframe (130-160 ms post-stimulus, or 63-93 ms after the second flash, see Figure 6-11), the SnPM analysis found that sources in the post-central gyrus of the superior parietal lobe (Brodmann area 5) were significantly more active for the MULTI difference wave compared to the UNI difference wave. The second-largest difference was found in the inferior parietal lobule (Brodmann area 40). Put another way, from 63-93 ms following the second flash,
there was significantly more activity related to the second flash in these areas when it was preceded by a multi-modal compared to a uni-modal context stimulus.

In the late timeframe (300-320 ms post-stimulus, or 233-253 ms after the second flash, see Figure 6-12), the SnPM analysis found that sources in the occipital lobes (Brodmann areas 18 and 17 had the highest and second-highest pseudo-$F$ values, respectively) were significantly more active for the MULTI difference wave compared to the UNI difference wave. In other words, from 233-253 ms following the second flash, there was significantly more activity in the primary and secondary visual cortices when the second flash was preceded by a multi-modal compared to uni-modal context stimulus. Table 6-3 shows the locations found by the SnPM analysis in both the early and late time intervals.

<table>
<thead>
<tr>
<th>Timeframe</th>
<th>Location of maximum difference</th>
<th>Location of secondary maximum difference</th>
<th>Pseudo-$F$ UNI ≠ MULTI</th>
<th>$p$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Early 130-160 ms</td>
<td>(-35, -50, -60 mm), BA 5: post-central gyrus, superior parietal lobe.</td>
<td>BA 40: Inferior parietal lobule, parietal lobe</td>
<td>2.11</td>
<td>.001</td>
</tr>
<tr>
<td>Late 300-320 ms</td>
<td>(10, -100, 15mm), BA 18: cuneus, occipital lobe</td>
<td>BA 17: cuneus, occipital lobe</td>
<td>1.80</td>
<td>.02</td>
</tr>
</tbody>
</table>

Table 6-3. Locations of the maximum difference (maximum pseudo-$F$ statistic) in the early and late time intervals.

6.6 Discussion

The experiment described in this chapter found that the electric field strength related to a flash stimulus was stronger when it was preceded by a multi-modal flash/beep stimulus, compared to when it was preceded by another uni-modal flash stimulus. This difference was found to be significant in two distinct timeframes – an early timeframe, from 130-160 ms, and a late timeframe, from 300-320 ms. Source localisation analysis found that the increased activity in the early interval was localised to a diffused area centred on the inferior and superior
parietal lobes, whereas the later increase was associated with stronger activity in a highly focussed area centred on primary and secondary visual cortex, in the occipital lobe.

The results indicated that processing of a visual stimulus was affected by the presence of an immediately prior multisensory event. Stronger neural activity related to a flash stimulus was found in the parietal and occipital lobes when the flash was preceded by a multi-modal context stimulus compared to a uni-modal context stimulus. The only difference between the uni-modal and multi-modal context stimuli was the presence or otherwise of a beep (see Figure 6-5). The uni-modal context stimulus consisted of a single flash, whereas the multi-modal context stimulus consisted of a simultaneous flash/beep. Relatively long-lasting interactions generated by the initial auditory and visual stimuli altered the processing of a subsequent visual stimulus.

### 6.6.1 A re-visit of experimental logic and relative timing

A matter of potential confusion in the interpretation of the results from the current study lies in the description of the timing of the effects. First, a brief re-visit of the logic of the analysis: The zero time point in the analysis was defined as the onset of the initial context stimulus, at which point either a flash (uni-modal context), or a combined flash/beep (multi-modal context) occurred. Approximately 67 ms after the uni/multi-modal context stimulus, another flash stimulus occurred (Figure 6-1). The response to either a single flash at 0 ms (for the UNI difference wave) or a combined flash/beep at 0 ms (for the MULTI difference wave), both with no second flash at 67 ms, was then subtracted, leaving behind responses to only the second flash (Figure 6-5). As each wave contained evoked neural responses only to the second flash, any difference between the two waves must have been due to a ‘lasting’ effect of the presentation of the different types of context stimulus before the second flash. The analysis subsequently found two such time intervals in which the UNI and MULTI difference waves were significantly different from each other. Responses to the second flash were larger when a multi-modal context stimulus was presented compared to a uni-modal
context stimulus. This increase in activity must necessarily have been caused by the multi-modal context stimulus, but was not directly evoked by the stimulus itself, as the neural activity evoked to the same flash/beep stimulus without the second flash was subtracted away. In discussing these interactions, it could be argued that timings should be measured in relation to the second flash only: as the initial flash/beep stimuli were presented simultaneously, and the timing of the second flash was always constant in relation to the initial context stimuli, the UNI and MULTI waves were essentially ERPs stimulus-locked to the second flash, with effectively a very long baseline extending to 100 ms before the initial context stimuli. However, despite this possible interpretation, the discussion to follow will consider the timings in relation to the first flash or flash/beep context stimulus, as it is these initial context stimuli which were responsible for the modulations of activity found, even though these changes were modulations of activity evoked by the second flash.

6.6.2 Involvement of the parietal lobe

In the current study, the effect of the multi-modal context stimulus was found in two main intervals, each in different regions of the brain. In the early interval (130-160 ms), increased activity in inferior and superior parietal lobes was found, and in the later interval (300-320 ms), increased activity was found in the primary and secondary visual cortex. The parietal lobes have traditionally been considered ‘association cortex,’ where information from separate sensory processing pathways is finally combined to form a unified sensory space (Andersen, 1997). Imaging studies in humans have found the area to be both multi-modal, or responsive to stimulation in more than one modality, as well as an area of integration, displaying non-linear super- or sub-additive response characteristics (Calvert, 2001, Calvert et al., 2001).

As was discussed in section 2.3.3.4 of Chapter 2, direct connections between the parietal and auditory cortex and the visual cortex have been found using tracer techniques in primates (Rockland and Ojima, 2003, Clavagnier et al., 2004, Falchier et al., 2002). More recently, intra-cranial recordings in awake
humans undergoing surgical planning procedures for intractable epilepsy have provided a timeline of the visual, auditory, and auditory-visual activity in these regions (Molholm et al., 2006). In this study, a speeded detection task using simple auditory, visual, and audio-visual stimuli was employed. Although no illusory stimuli were presented, the short combined flash/beep stimulus was very similar to the multi-modal context stimulus used in the current study. The accuracy and reaction times to the audio-visual stimulus in Molholm et al (2006) showed the expected violation of the race model, indicating that the facilitation of behavioural responses was not simply due to the summation of probabilities of responses for the two uni-modal stimuli alone (see discussion section of Chapter 3 for an explanation of the race model). Grid electrodes over the parietal cortex in three patients showed the onset of neural responses to the auditory stimuli at around 30 ms, and the visual stimuli at around 75 ms. Non-linear responses to the audio-visual stimulus (where the response to the audio-visual stimuli was significantly different than the sum of responses to the auditory and visual stimulus alone) were found in exactly the same locations with onset from between 120-160 ms.

Together with the neuroimaging data (Calvert, 2001, Calvert et al., 2001), the study by Molhom et al (2006) strongly suggests that the parietal lobes are sites of multisensory integration, and not only co-activation. The results also provide a timeline of activation in the parietal lobes: auditory activation arrives first, at around 30 ms post-stimulus, followed by visual activation, at around 75 ms. Auditory and visual stimuli are then integrated, with non-linear responses in the same area from around 120 ms. Interestingly, increased MULTI responses in the early interval from the current study were also found in the parietal lobes, in a very similar interval to that found by Molholm et al (2006). In the MULTI difference wave, an initial flash/beep stimulus was followed 67 ms later by another flash stimulus, and subsequently the responses to a second flash/beep stimulus were subtracted. The initial flash/beep multi-modal context stimulus in the current study was very similar to the multi-modal stimulus used in Molholm et al (2006), suggesting the possibility that the increase in activity found in the current study
was also indicative of multi-sensory processes in the parietal lobes, driven by the combined audio-visual context stimulus.

6.6.3 Timing of interactions found in other flash-beep studies

The fact that the timing of the stronger responses to flash stimuli in multi-modal contexts found in the current study aligned well with the location and timing of multi-sensory interactions in the parietal lobes suggests that the initial multi-modal flash/beep stimulus affected the subsequent processing of another visual stimulus by engaging multi-sensory processing pathways. Figure 6-13 shows a timeline of the first neural responses in A1, V1, and the parietal lobes (PL), as well as multi-sensory interactions found from various flash-beep illusion studies to combined flash/beep stimuli. Figure 6-13 is complex and each part will be discussed in turn. Before discussion of the illusion studies, an overview of the networks engaged by the presentation of a single simultaneous flash/beep stimulus is in order.
Figure 6-13. Timing diagram showing various responses to a simultaneous flash/beep stimulus at time zero. Some points to be identified further below are in response to a difference between stimuli, in which case there is also a flash at 67 ms (for the current study), or a beep at 67 ms (Shams et al 2001, 2005, Mishra et al 2007), in which case there is also an ‘illusory flash’ at approximately 120-130 ms. The green and red circles of points 1 and 2 show the times at which activity to auditory (green circles) and visual (red circles) stimuli first appears in the primary auditory (A1) and visual (V1) areas. Point 3 shows when the activity from the second flash, isolated by the MULTI waveform (2F1B – 1F1B), could be expected to occur. Note that in this case, auditory activity from the initial flash/beep is subtracted and would not appear, leaving only point 3. The green and red inverted triangles (points 4, 5 and 6) show when uni-modal activity from a combined flash/beep at 0 ms reaches association areas in the parietal lobes (PL - 4,5), and when super-additive responses to audio-visual stimuli are found in the same areas shortly afterwards (mixed green and red inverted triangles, 6). The blue boxes show the timing of increased activity found in the current study: points 7, 8: ‘early’ and ‘late’ intervals, respectively, at which MULTI response was significantly larger than the UNI response in the current...
study. Points 9 and 10 (pink triangles) show increased activity in V1 in participants who were pre-disposed to illusion (point 9), and increased negativity in A1, in illusion compared to non-illusion trials (point 10), in Mishra et al (2007). Point 11 shows the decreased GFP localised to V1 in MULTI compared to UNI responses in Meylan and Murray, (2007). Points 12 & 13 show interactions (1F2B > 1F0B+0F2B) found in parietal (12) and occipital (13) MEG sensors in Shams et al (2005a). Points 14 & 15 show significant interactions (1F2B > 1F0B+0F2B) at occipital EEG electrodes in Shams et al (2001). Times for points 1-3 were extrapolated from primate electrophysiology data to human-equivalent times using ‘3/5ths rule’ (Musacchia and Schroeder, 2009). Details: 1: first onset to broadband noise stimulus in macaque A1 (Lakatos et al., 2005). 2, 3: response to first and second light flash in macaque V1 (Chen et al., 2007). 4, 5, 6: first onset time of activity to auditory (4), visual (5), and integrative response to audio-visual stimulus (6) in human parietal lobe (Molholm et al., 2006). Note there are some small timing differences between the paradigms in Meylan & Murrray (2007), who used an ISI of 52 ms; Mishra et al (2007), 70 ms; Shams et al (2005a, 2001), 67 ms; and the current study, also 67 ms. Full paradigm details in Table 3-1 in Chapter 3.

The simplest case is when a single beep is paired with a flash. This was the stimulus configuration in Molholm et al (2006). The stimulus used in that study is very similar to the multi-modal context stimulus in the current study, and this combined flash/beep stimulus was also present for each of the illusion-capable stimuli in the current study as well as others - 1F2B and 2F1B stimuli both contain a combined flash/beep stimulus at 0 ms. In this simplest case, the green and red dots on the diagram (points 1 and 2) show the initial activation of auditory and visual primary sensory cortex, respectively. As can be seen, the first activation of auditory cortex is considerably faster, around 10-14 ms (Lakatos et al., 2005), than the first activation of visual cortex, around 50 ms (Chen and Yeh, 2009). Although the studies of first neural responses in primary auditory and visual areas were
performed using multi-unit activity studies in primates, the timing for this discussion has been extrapolated to human-equivalent times using the ‘3/5ths rule’ (Musacchia and Schroeder, 2009, Schroeder et al., 2004). Rapidly following each of these primary sensory activations, there is activation of the parietal lobe (green and red inverted triangles, Molholm et al., 2006), firstly by the auditory stimulus, at about 30 ms (point 4), and then by the visual stimulus at around 70 ms (point 5). Feed-forward connections from primary sensory areas to parietal areas are shown with dotted blue arrows (Jones and Powell, 1980). As the auditory and visual stimuli were presented together, there is also a non-linear interaction (overlayed green and red inverted triangles, point 6) in the parietal lobe, beginning from around 120-160 ms (Molholm et al., 2006). Note that it is unknown how long this parietal activation persists, only that the onset is from 120-160 ms. Subsequent feed-back connections from the parietal lobes back to primary sensory areas are shown with dotted red arrows (Rockland and Ojima, 2003). As was reviewed in Chapter 2, the parietal lobes are known to have top-down influences on multisensory processes in other cortical areas, such as the primary auditory and visual areas, as well as in subcortical areas such as the SC. Direct cortico-cortical connections between A1 and V1 are shown with solid brown arrows (Clavagnier et al., 2004, Falchier et al., 2002). In summary, the presentation of a simultaneous flash/beep stimulus activates the auditory cortex, the visual cortex, and the parietal cortex (and other regions not discussed here). Multi-sensory interactions occur in the parietal cortex beginning from approximately 120-160 ms after the presentation of the multi-modal stimulus. Feedback connections may be active from parietal areas to primary sensory areas, as well as between primary sensory areas.

Although responses to the fission illusion stimuli were not tested in this study, the fission illusion stimulus (1F2B, or a combined flash/beep followed by another beep) has an onset with exactly the same characteristics (a combined flash/beep) as the multi-modal context stimulus in the current study, and similar characteristics to the multi-modal stimulus in Molholm et al (2006). As single beeps presented in isolation are not known to elicit flash illusions, the effect of the
multi-modal stimulus preceding the second beep in the fission illusion stimulus is likely to be involved in triggering the second beep to elicit the flash sensation, in the same way that the multi-modal stimuli in the present study affected subsequent processing of the second flash. The findings suggest three possible mechanisms by which the illusion may operate: either through direct cortico-cortical connections between A1 and V1, through feedback connections from the parietal cortex, or by some combination of the two. Figure 6-13 shows hypothetical activation of the auditory cortex by the second beep in the 1F2B illusion stimulus (green circle at 80 ms) and the approximate time at which the illusory flash might be expected to occur (unfilled red circle, point 3). As the timing of the perception of the illusory flash has not been experimentally determined, this position is speculative, and is based on the timing of activation reaching V1 from a second ‘real flash,’ if one were to occur. However, it is quite possible that the perception of the illusory flash is either later (likely) or earlier (unlikely) than this point, and the timing may vary between individuals. Given this caveat, however, it is interesting to compare this proposed timing of the illusory flash with the multi-sensory interactions found in the parietal lobe elicited by the initial flash/beep stimulus (Figure 6-13 point 6, Molholm et al., 2006), as well as the results from the current study, where increased activation of the parietal lobes was found in isolated responses to the second flash at the same time (Figure 6-13, point 7). The fact that ERP responses to the second flash were modulated by the presence of the initial multi-modal context stimulus, in the same timeframe and in the same brain areas where multi-sensory interactions to the initial flash/beep context stimulus were found, strongly suggests that processing of the second flash was affected by long-lasting multi-sensory interactions.

---

5 Note that for the stimuli that contain two flashes, the original timing of the second real flash in Shams et al (Shams, L., Kamitani, Y. & Shimojo, S. 2000. Illusions. What you see is what you hear. *Nature*, 408, 788, Shams, L., Kamitani, Y. & Shimojo, S. 2002. Visual illusion induced by sound. *Brain Research. Cognitive Brain Research*, 14, 147-52) was determined by pilot testing in which several participants were asked to match the presentation time of a flash stimulus with the perceived interval to the illusory flash (personal communication with Laden Shams, April 2003).
processes in the parietal lobes that were most likely engaged by the initial multi-modal context stimulus.

### 6.6.4 Illusions and involvement of the occipital lobe

The results reviewed above are in relation to the effect of a single multi-modal stimulus on subsequent uni-sensory processing. By comparing the responses to 1F2B stimuli to the sum of 1F0B and 0F2B responses, a number of flash-beep illusion studies have also found non-linear interactions related to the perception of the illusory second flash in the same ~110-160 ms timeframe and in similar brain regions. This interaction waveform is commonly referred to in the literature, and is usually written as AV- (A + V). Table 6-4 shows the details of these studies, and selected results are also depicted using coloured rectangles in Figure 6-13. For instance, the MEG study of Shams et al (2005a) found illusion stimulus interactions at both parietal and occipital MEG sensors in the same 120-160 ms timeframe as the multi-sensory interactions from Molholm et al (2006) and the current study found in the parietal lobes only (Fig 6-13, points 12 and 13). In this case, the interactions were linked to perception of the illusory flash in response to the 1F2B illusion stimulus, and were significant when the response to the 1F2B stimulus was greater than the sum of 1F0B and 0F2B responses. Mishra et al (2007) also found interactions in the same timeframe. An increase of ERP activity localised to V1 just prior to the expected time of the illusory flash was found by these authors, but only in a subset of participants who were pre-disposed to the illusion (Figure 6-13, point 9). Prior again to the activity in V1, an enhanced negativity localised to A1 was also found, this time in trials where the illusion actually occurred compared to those where it did not occur (Figure 6-13, point 10). While activation of V1 (30-60 ms after the second beep) was necessary but not sufficient for perceiving the illusory flash, the earlier activation of A1 (only 20-40 ms after the second beep) in illusion trials only appeared to be the obligatory trigger for the illusion. The authors proposed that rapid interplay at least partly via direct cortico-cortical connections between A1 and V1 was responsible for perception of the illusory flash.
Shams et al (2001) investigated the effect of auditory stimuli on the flash visual-evoked potential (VEP), using a slightly altered version of their original flash-beep paradigm. ERPs were examined by comparing the 1F2B responses to the sum of the flash alone (1F0B) and beeps alone (0F2B) stimuli. The amplitude of each point in the grand average waveform was then compared against zero. If more than 15 consecutive points (15 ms) in the interaction waveform were significantly different from zero, a multisensory interaction was deemed to have occurred. When the stimuli were presented in the visual periphery, an ‘early’ interaction was found at occipital electrodes between 170 and 200 ms, and a ‘late’ interaction from 260 to 360 ms. The early time interval corresponded to ~90-140 ms after the second beep – the stimulus that is presumably responsible for the generation of the illusion. The authors therefore interpreted this effect as an indication that the second beep ‘activated’ the primary visual cortex, thus generating the percept of the illusory second flash.

As well as the early differences in the parietal lobes, stronger activity related to the second flash when it was preceded by a multi-modal context stimulus was also found in the current study in the occipital lobes, at the relatively late time of 300-320 ms. This time and location corresponds with the finding of ‘late’ interactions (1F2B > (1F0B + 0F2B)) at occipital electrodes in Shams et al (2001), as well as the decreased activity in MULTI compared to UNI difference waves in Meylan and Murray (2007). Shams et al (2001) compared the 1F2B > (1F0B + 0F2B) interaction waveform with activity at the same electrode evoked by a real flash (ie 2F – 1F), and found that the waveforms were indistinguishable. They therefore concluded that the interaction waveform reflected modulation of visual-specific processing by the auditory stimulus. It is also interesting to compare the current results with those of Watkins et al (2006), who found that fMRI activation in retinotopically-mapped V1 was stronger for 1F2B stimuli on trials when illusory flashes were reported compared to when the stimulus was perceived veridically. As has been previously mentioned, the 1F2B stimulus contains the same initial multi-modal flash/beep stimulus that was considered as the ‘context’ stimulus in the current study. In the current study, it was followed by another flash stimulus, and
the activity related to another flash/beep stimulus, but without any following second flash, was then subtracted, leaving only the response to the second flash. To follow the analogue, in Watkins et al (2006) the multi-modal flash/beep stimulus was followed by another flash, and it was found that activity in V1 increased when this extra flash also caused the perception of an additional beep. As with the current study, the modulation of visual processing that led to increased activity in V1 in Watkins et al (2006) may have been caused either by direct cortico-cortical connections from A1 to V1, or by the modulation of V1 by multisensory processes set in motion by the initial multi-modal flash beep stimulus.

Overall, incorporating the results from the current study with previous studies showing multisensory interactions in the parietal lobe following a combined flash/beep stimulus suggests an involvement of both feedback from parietal to primary sensory areas, as well as direct connections between primary sensory areas.
<table>
<thead>
<tr>
<th>Study</th>
<th>Description of methods</th>
<th>Timing and location of effects</th>
</tr>
</thead>
<tbody>
<tr>
<td>Shams et al 2001</td>
<td>ERP interaction waveform: 1F2B - (1F+2B) significantly different from zero.</td>
<td>Significant intervals at O1, Oz and O2 electrodes, around 170-200 and 250-300 ms. Equivalent to ~100-130 and 180-230 ms after second beep.</td>
</tr>
<tr>
<td>Shams et al 2005</td>
<td>MEG interaction waveform: 1F2B – (1F+2B) significantly different from zero.</td>
<td>1) Earliest interaction at 1 occipital sensor from 35-65 ms (15 ms before the second beep)! 2) More robust interaction at occipital, temporal, parietal, and frontal sensors from 125-170 ms (75-120 ms after second beep). 3) Widespread interactions across most sensors from 340-450 ms (290-400 ms after second beep).</td>
</tr>
<tr>
<td>Mishra et al 2007</td>
<td>Difference ERP isolating response to illusory second flash.</td>
<td>1) In Ss pre-disposed to illusion: Increase of activity localised to V1 from 30-60 ms after second beep. 2) In illusion trials compared to no-illusion trials: Increased activity localised to auditory cortex from 20-40 ms after second beep. 3) Following these early effects, a complex series of interactions in visual cortex and superior temporal cortex.</td>
</tr>
<tr>
<td>Meylan &amp; Murray 2007</td>
<td>Difference ERP isolating response to second flash with uni-modal vs multi-modal context</td>
<td>GFP lower from 255-297 ms (160-232 ms following second flash) when a multi-modal context stimulus preceded rather than uni-modal context stimulus. Difference localised to left posterior occipital lobe, BA 18.</td>
</tr>
</tbody>
</table>

Table 6-4. Details of flash-beep illusion studies finding interactions in the parietal lobes, A1 and V1.

### 6.6.5 Comparison to Meylan & Murray (2007)

The present study was similar in design to Meylan and Murray (2007). However, there are several important differences in the results. Firstly, there were small differences in the behavioural results – while Meylan and Murray (2007) report very high accuracy rates (M = 89%) for the fusion (2F1B) stimulus, accuracy rates were slightly lower in the current study (M = 77.2%, SD = 29.1), indicating the presence of fusion illusions on a small number of trials in the current study. However in both studies, only correctly-responded trials were analysed. In other respects, the behavioural results were very similar, with accuracy above 90% for all non-illusion stimuli in both studies. For the fission (1F2B) stimulus, mean accuracy was close to 50% in both studies, a figure consistent with
the vast majority of flash-beep illusion studies (see Table 3-1 in Chapter 3 for a summary of a number of flash-beep illusion methods and results).

Meylan and Murray (2007) found only one time interval in which the electric field response for the UNI and MULTI waves (also measured using the GFP) was different. This interval was from 238-275 ms, directly in-between the two significant intervals found in the present study. The other major difference lies in the direction of the result. Whereas Meylan and Murray found the MULTI GFP waveform to be of overall lower amplitude than the UNI GFP waveform, and significantly lower in the aforementioned time interval, the MULTI GFP waveform was consistently larger than the UNI waveform in the current study.

The differences between the two results may be due either to participant factors or to experimental methods/analysis factors. The age ranges were very similar between the two studies, although while the sample in the current study was predominantly female (8 females, 4 males), the sample in Meylan & Murray (2007) was predominantly male (6 males, 2 females). However, accuracy scores for any stimulus in the current sample did not differ significantly for males and females, making the gender distribution an unlikely source of the difference in results. Similarly, there were small differences in the stimulus timing used. Table 6-5 shows stimulus details for both studies. While the durations of the auditory and visual stimuli are longer in the current study, the onset times of the auditory and visual stimuli are identical. Differences in the stimulus durations are unlikely to account for the differences in the neurophysiological results, although it is possible that the longer flash duration (and subsequent reduction in the blank inter-stimulus interval between flashes) in the current study may have rendered the visual double-flash stimulus more likely to ‘fuse’ into a single flash, thus accounting for the difference in accuracy scores for the 2F1B trials. Participants in the current study were instructed to ‘count the number of flashes that they saw while ignoring the beeping sounds.’ The exact instructions are not reported in Meylan and Murray (2007), however the procedure section states that ‘Subjects’ task was to indicate the number of flashes perceived via a serial response box.’
<table>
<thead>
<tr>
<th>Study</th>
<th>SOA (ms)</th>
<th>Refresh rate (Hz)</th>
<th>Flash Duration (Hz, refresh periods)</th>
<th>Flash location</th>
<th>Flash size</th>
<th>Beep duration</th>
</tr>
</thead>
<tbody>
<tr>
<td>Current study</td>
<td>66.7 ms</td>
<td>75 Hz</td>
<td>26.7 ms, 2 periods</td>
<td>7.5° below fixation</td>
<td>3°</td>
<td>8 ms</td>
</tr>
<tr>
<td>Meylan &amp; Murray 2007</td>
<td>66.7 ms</td>
<td>75 Hz</td>
<td>13.3 ms, 1 period</td>
<td>6.3° below fixation</td>
<td>2.2°</td>
<td>13 ms</td>
</tr>
</tbody>
</table>

Table 6-5. Stimulus configuration differences between the current study and Meylan & Murray (2007).

Differences in the EEG recording setup and analysis are also unlikely to account for the differences in results. Although Meylan and Murray (2007) used 128 recording sites rather than the 60 used in the current study, the increase in the number of electrodes is more likely to affect source localisation accuracy rather than the GFP measure. In many other respects the analysis of EEG data was similar – the average reference was used, the UNI and MULTI subtractions were performed in the same way, and the GFP was calculated in the same way. In short, it is difficult to explain the large differences in result between the two studies. Individual differences and the composition of the samples in each case may explain the differences in results.

6.6.6 Comment on the very early and mid-latency differences in the ERP t-intensity plot that were non-significant in the GFP tests

From the scalp ERP plots and the t-intensity surface plot in Figure 6-8, it is apparent that there were several groups of electrodes with significant differences in a very early interval, between 0-40 ms following the first flash, and in a ‘middle’ interval between the early and late intervals analysed, between 210-240 ms. However, these differences did not show up when testing the difference between the GFP for UNI and MULTI stimuli within this time period. In the case of the very early interval, the differences between the UNI and MULTI waves appeared in both the ERP data as well as the GFP data, although the length of the interval...
(3 sample points, approximately 1.5 ms) did not survive the auto-correlation threshold of 10 (20 ms) sample points, and so was not sufficiently long to be deemed reliable (Guthrie and Buchwald, 1991).

As can be seen from the intensity plot, there was a cluster of electrodes centred at CPz in the ‘middle’ interval (centred at around 210 ms) where the MULTI response was stronger than the UNI response, and surrounded by electrodes to the left and right where the opposite was the case (MULTI < UNI). This pattern was repeated with a cluster of electrodes centred around Pz, where the MULTI response was again stronger than the UNI response, with surrounding electrodes to the left and right where the opposite was the case. In the GFP plot in the same period, however, there were no significant differences in the overall response strength between the UNI and MULTI conditions, and so the potential sources of these clusters of significant differences seen in the ERP intensity plot were not investigated. The reason these difference did not appear the GFP waveform is due to the nature of the GFP measure. The GFP is calculated at each time point by taking the standard deviation of the voltage values at every electrode. This is equivalent to calculating the sum of squared potential differences between all electrode combinations, and represents a measure of overall electric field strength across the entire electrode array. As can be seen in the plot of the GFP over time, the GFP for both the UNI and MULTI GFP drops sharply at approximately 220 ms, before both increase together during the non-significant interval.

In periods where the GFP is low, the associated topographic maps are less organised, and are characterised by patchwork patterns where adjacent areas on the map are strongly positive or negative. These transition periods are thought to separate ‘microstates’ of the brain (Lehmann and Skrandies, 1980, Michel et al., 1993), where a particular stable set of brain generators/functions are giving way to a new set. This can be illustrated with a series of topographic maps across the periods involved. Figure 6-14 shows the GFP for UNI and MULTI conditions, with the early and late timeframes marked on the x-axis. Below are a series of topographic maps for each condition, with one map every 20 ms. As can be seen,
there are periods of map stability which coincide with periods where the overall response strength (the GFP) is high. In the 220-250 ms interval, when the GFP drop sharply, especially for the MULTI condition, the maps can be seen to become more chaotic, before settling into a new pattern.

Figure 6.14. A series of topographic maps spanning the timeframes in which there was a significant difference between the GFP for the UNI (top row) and MULTI conditions (indicated by the blue bars on the x-axis), as the non-significant period in-between. Each map was calculated by taking the average over a 20 ms interval.

6.7 Conclusion

Overall, the results from the current experiment showed that neural responses related to processing of a visual stimulus could be modulated by prior multisensory stimuli. Specifically, EEG responses to a flash stimulus were stronger when a flash stimulus was immediately preceded by a multi-modal flash-beep stimulus compared to when it was preceded by another uni-modal flash. The differences were localised to the superior and inferior parietal cortex from 130-160 ms, and to the primary and secondary visual cortex from 300-320 ms. The results are supportive of views implicating the involvement of higher-order multi-sensory
association regions in uni-sensory processing, but cannot also rule out the involvement of direct cortico-cortical connections between primary sensory areas.
7 Children are more susceptible than adults to the flash-beep illusion


Reproduced exactly as published, apart from figure/heading labels and other minor formatting changes as required, with permission from John Wiley and Sons, who retain copyright.
7.1 Introduction

In the mature human brain, the process of multisensory integration (MSI) can result in significant perceptual advantages, particularly for stimuli where the temporal and spatial properties in each modality are congruent. For example, the detection of weak visual stimuli can be enhanced by concurrently presented sounds (Frassinetti et al., 2002a, Lovelace et al., 2003) and the intelligibility of speech is improved when video of the appropriate lip movements is available (Sumby and Pollack, 1954, MacLeod and Summerfield, 1987). This integration process generally improves perceptual estimates. However, if incongruent stimuli are presented to multiple sensory modalities, integration can still occur, resulting in altered or illusory percepts. In the ventriloquist illusion, for example, the perceived location of sound sources can be strongly influenced by spatial visual information (Pick et al., 1969, Welch and Warren, 1980), and video of incongruent mouth movements can affect the perception of speech (McDonald et al., 2003, McGurk and MacDonald, 1976).

The ability to match or compare information from multiple sensory signals appears quite early in human development. For example, infants can match the intensities of auditory and visual signals (Lewkowicz and Turkewitz, 1980), discriminate between redundantly-specified auditory-visual tempos (Bahrick et al., 2002), and at only 1-3 days post-natal are able to match non-native primate vocalisations to the appropriate visual gesture (Lewkowicz et al., 2010). However, as with many unisensory abilities (such as speech perception (Kuhl et al., 2003), face perception (Pascalis et al., 2002), and music perception (Hannon and Trehub, 2005)), some multisensory abilities present at birth or soon after show evidence of narrowing or tuning with post-natal development. For example, the ability to respond to an audio-visual primate phonetic contrast has been shown to narrow between 6 and 11 months of age (Pons et al., 2009). This finding and many others (for review see Lewkowicz and Ghazanfar, 2009) suggests that post-natal experience with the appropriate stimulus-response contingencies is required in order to tune the matching processes to best suit the child’s environment.
Multisensory abilities such as intensity matching (Lewkowicz and Turkewitz, 1980), audio-visual rhythm discrimination (Bahrick and Lickliter, 2004, Bahrick and Lickliter, 2000) and the detection of temporal equivalence in the form of speech onsets and offsets (Lewkowicz, 2010, Lewkowicz et al., 2010), can appear very early in infancy. Although these abilities appear to be present at birth (and are often subsequently refined or narrowed with post-natal experience), other multisensory phenomena emerge later in life. For instance, in the stream-bounce illusion (Sekuler et al., 1997), the presentation of a brief click as two moving balls intersect on a video display biases the perception towards bouncing rather than streaming in adults. There is evidence from looking-time measures that infants of 6 and 8 months of age, but not 4 months of age, spend more time looking at the “streaming” display, where the sound is absent, than at the “bouncing” percept to which they are habituated (Scheier et al., 2003a). The looking-time measures suggest that the streaming display, where the only difference is the lack of the click sound at the collision point, is visually novel. Although there is some controversy over the interpretation of looking time measures (Slater, 2003 for discussion, see Scheier et al., 2003b), the evidence suggests that the illusion may not be perceived until at least 6 months of age. Similarly, Neil, Chee-Ruiter, Scheier, Lewkowicz and Shimojo (2006), found that spatial orienting responses to auditory-visual targets were facilitated in 8-10 month-old infants as well as adults, but not reliably in infants under 8 months of age.

Although the McGurk illusion has been found to occur in infants (Rosenblum et al., 1997), incongruent visual stimuli have less of an effect on final phoneme perception in both 3-5 and 7-8 year-old children than in adults (McGurk and MacDonald, 1976, Massaro, 1984, Massaro et al., 1986). In both the McGurk and stream-bounce illusion cases, information from one sense alters or enhances perception in the other, requiring a more integrative process than tasks requiring the transfer or matching of information across the senses. Gori, Del Viva, Sandini, and Burr (2008) have likewise shown that judgements of size and orientation do not benefit from the ability to optimally integrate visual and haptic information until between 8 and 10 years of age, and Nardini, Jones, Bedford, and Braddick
(2008) have shown that children up to 8 years of age do not integrate self-generated motion cues with visual cues in an object navigation task. Barutchu, Danaher, Crewther, Innes-Brown, Shivdasani and Paolini (2010), using a simple auditory-visual detection task, also found that reaction times to auditory-visual stimuli do not show mature super-additive enhancement in children as old as 11 years of age. Using a similar detection task, Brandwein, Foxe, Russo, Altschuler, Gomes and Molholm (2010) have recently shown that mature levels of multisensory facilitation (and associated auditory event-related potentials) are only reached by approximately 15 years of age. In summary, although infants can match information across the senses at a very young age, the facilitation of perceptual judgements and reaction times by the integration of multisensory stimuli is generally found to develop slowly throughout childhood.

Another multisensory illusion is the flash-beep illusion. In this illusion, the presence of a number of brief auditory stimuli affects judgments of the number of visual flashes presented (Shams et al., 2000). When a single flash is paired with two or more beeps, participants often report seeing more than the single flash that was presented. This form of the illusion has been termed a fission illusion, as the double auditory stimulus is thought to split the perception of a single flash into two. The fission illusion has since been shown to be robust to a degree of temporal delay between the auditory and visual stimuli (Shams et al., 2002), spatial separation of the auditory and visual stimuli across the visual midline (Innes-Brown and Crewther, 2009), and even accuracy feedback on each trial specifically designed to reduce its strength (Rosenthal et al., 2009).

Several neuro-imaging studies (Bhattacharya et al., 2002, Mishra et al., 2007, Shams et al., 2001, Watkins et al., 2006) have shown that reports of the fission illusion are correlated with increased activity in primary visual cortex. These studies generally support the hypothesis that the illusion results from the integration of auditory and visual information, rather than the possible introduction of response biases. In a related sound-induced fusion illusion, a single beep causes the fusion of a double flash into the perception of a single flash. Neural correlates of the fusion illusion, measured using functional magnetic
resonance imaging (fMRI - Watkins et al., 2007) and event-related potentials (ERP - Mishra et al., 2007), are correspondingly reduced in the same areas. In combination with the behavioural evidence, these findings of increased activity in V1 during perception of illusory extra flashes (fission) and reduced activity in V1 during illusory reduction of flashes (fusion) strongly suggest that the illusion occurs as a result of modification of activity in primary visual cortex by the auditory stimulus. Thus, reports of illusory fission or fusion of a visual stimulus by sound are indicators of integration at low levels in the brain, possibly in the primary visual cortex.

The flash-beep paradigm is simple to administer and suitable for assessing MSI in both adults and children. However, to our knowledge, only one study has investigated the flash-beep illusion in children (Tremblay et al., 2007). In this study, both the fission and fusion illusions were studied in children between 5-19 years of age, separated into three age groups. Both fission and fusion illusions were shown to occur in each age group; however, no significant differences in the number of illusions for either fission or fusion stimuli were found between any of the three age groups. As there was no gradual improvement or decline in accuracy with age, the authors interpret their results as consistent with theories suggesting that audio-visual integration abilities mature very early in life. However, illusions were reported more often (about 80% illusion responses in fission illusion trials) than previous studies have found in adults – generally around 50% illusion responses in fission illusion trials – (Andersen et al., 2004, Shams et al., 2000, Shams et al., 2002, Courtney et al., 2007, Meylan and Murray, 2007, Innes-Brown and Crewther, 2009). Given this difference, it is possible that the form of multisensory integration measured by the flash-beep illusion task is actually immature, and over-active in children, with auditory information being more often, but less selectively integrated with the visual signal. Without an adult control group completing the same paradigm, this possibility is difficult to assess.

In the present study, the objectives were to determine whether children showed less selective integration of auditory and visual temporal information than adults. Using the flash-beep paradigm to provide evidence of multisensory
integration, reports of fission and fusion illusions by normally developing children were compared with adults. Six combinations of single and double flash/beep pairs were presented, including “fission” and “fusion” illusion stimuli where the numbers of flashes and beeps did not match. As the performance of children in the flash-beep task is still relatively unknown, multisensory control trials using congruent flash/beep stimuli were also presented, and unisensory visual-alone trials were presented to test whether the visual stimulus itself was unambiguous for both adults and children. Accuracy and reaction time measures were compared for adults and children with each stimulus type to determine the extent to which MSI caused incorrect reports of the flash stimuli. To determine any developmental trajectory, these measures were correlated with age within each group.

7.2 Methods

7.2.1 Participants

Fifty-two individuals participated in the study. Thirty participants under the age of 18 were assigned to the child group, and twenty-two over the age of 18 were assigned to the adult group. The child group consisted of 14 girls and 16 boys, with a mean age of 11.1 years (SD=2.0, range 8-17 years). Children were recruited from schools in the Melbourne metropolitan area under the auspices of the Catholic Education Office, Melbourne, and came from a wide variety of socio-economic circumstances and cultural backgrounds. The adult group consisted of 12 females and 10 males with a mean age of 29.0 years (SD=5.8, range 19-42 years). All adult participants gave written informed consent, as did the parents of child participants. The study was approved by the Human Research Ethics Committee.

6 The oldest child was 17.0 years old, and the youngest adult was 19.0 years old. As these two participants’ ages fell close to the boundary dividing the adult and child groups, all analyses were repeated with these two individuals removed. No differences were found in the pattern of statistical significance across any of the analyses performed, so these individuals were included in the results.
Committee of the Royal Victorian Eye and Ear Hospital and by the Catholic Education Office, Melbourne. All participants were right-handed, had normal hearing, normal or corrected-to-normal vision, and reported no diagnosis of a psychological or neurological disorder.

### 7.2.2 Stimuli and apparatus

The experiment was conducted in a quiet sound-treated room with a background sound level of approximately 39 dB SPL (A-weighted). The visual stimulus consisted of a white disk, which flashed once or twice on a 20-inch cathode ray tube monitor (CRT) with a black background. The disk subtended 3° of visual angle and was located 7.5° below a fixation cross, which was positioned 2.5° above the centre of the screen. The refresh rate of the CRT was set to 160 Hz, and each flash consisted of two refresh periods (12.5 ms, see Figure 1). On selected trials, short beeps were presented, along with the flashes, from a speaker placed centrally under the CRT. The beep was a 3500 Hz, 83 dB SPL (A-weighted) sine wave of 8 ms duration, with 3 ms rise and fall times. The fixation cross remained present throughout each block. Participants sat in a comfortable armchair with the head 100 cm from the CRT. A response box rested on a desk at a comfortable distance directly in front of the CRT. The auditory and visual stimuli were generated by programmable real-time hardware (RX6 processor, Tucker-Davis Technologies, Alachua, FL; Visage, Cambridge Research Systems, Kent, England), which ensured that each beep onset could be exactly synchronised with the required refresh of the computer monitor.

All participants completed the Purdue Pegboard (Tiffin and Asher, 1948), a measure of manual dexterity where participants were asked to insert the maximum number of small pins into a row of sockets in a given time. The task was run using each hand alone (Pegboard-Left, Pegboard-Right) and with both hands (Pegboard-Both).
### 7.2.3 Procedure

The procedure broadly followed that outlined in the original report of the sound-induced flash phenomenon (Shams et al., 2002), except that a slightly reduced number of stimulus combinations was presented due to the time constraints imposed by testing young children, and a simplified stimulus timing scheme was used. Whereas Shams et al. (2002) presented the first beep 23 ms prior to the first flash, the onsets of each auditory and visual stimulus were exactly synchronised in the current procedure. This stimulus timing is simpler to interpret than that used in the original report, and has been used successfully in fMRI (Watkins et al., 2006) and ERP (Meylan and Murray, 2007) investigations of the flash-beep illusion.

On each trial, there were either one or two flashes, along with zero, one, or two beeps. Trials are henceforth referred to using a code indicating the number of flashes followed by the number of beeps – “2F2B” thus refers to a trial with two flashes and two beeps. In trials with multiple flashes or beeps, the time between the onsets of successive beeps or flashes was 75 ms, corresponding to 12 CRT refresh periods. An example of a 2F2B trial is shown in Figure 1.

![Flash beep stimulus timing. Shown is an example of a 2-flash, 2-beep (2F2B) trial.](image)

Each trial began with the visual stimulus presented in the centre of the screen, below the fixation cross. Following each trial was a 2-3 s interval, during which the participant was required to indicate how many flashes were perceived by pressing one of two buttons on the response box. Participants were instructed to keep their gaze on the fixation cross during the whole trial block and to count the number of flashes that would appear while ignoring the beeps. They were
instructed to press a button labelled either “1” or “2” depending on how many flashes they saw, and to make their best guess if unsure.

The six possible flash/beep combinations (1F0B, 1F1B, 1F2B, 2F0B, 2F1B, 2F2B) were randomly presented ten times each in a single block. This block was repeated six times (with trials re-randomised each time). Each stimulus was thus presented 60 times. Button-press responses and reaction times (RT) were recorded for each stimulus presentation. Responses were not analysed if they were made less than 100 ms or more than 1500ms post-stimulus. Each block ran for an average duration of two minutes and breaks could be taken between each block. The total testing time was approximately 20 minutes.

7.3 Analyses and Results

7.3.1 Purdue Pegboard

All tests of statistical significance are reported as significant if $p < .05$. As reaction times form an important part of the later analyses, it was important to obtain a baseline measure of the manual dexterity of all participants. Table 1 shows the mean scores (number of pegs inserted in 30 seconds) for all the pegboard sub-scales in each group. Since all participants were right-handed, and responded with their right hands, only the Pegboard-Right scores were included in later correlation analyses. Adults were significantly faster than children when using the left, right, or both hands (see Table 1). There were no significant differences between males and females in either group for any of the Pegboard measures.
Table 7-1. Means, standard errors of the mean, and t-test outcomes for Purdue Pegboard Scores.

<table>
<thead>
<tr>
<th></th>
<th>Adults</th>
<th>Children</th>
<th>t-test</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pegboard-Right</td>
<td>16.0 (2.3)</td>
<td>13.2 (2.1)</td>
<td>t(46)=4.5 *</td>
</tr>
<tr>
<td>Pegboard-Left</td>
<td>15.1 (1.5)</td>
<td>12.7 (2.0)</td>
<td>t(46)=4.7 *</td>
</tr>
<tr>
<td>Pegboard-Both</td>
<td>12.9 (1.4)</td>
<td>10.8 (1.5)</td>
<td>t(46)=4.9 *</td>
</tr>
</tbody>
</table>

*Note. Pegboard scores are the number of pegs inserted in 30 seconds; Standard errors of the mean are given in parentheses; *p < .05 (two-tailed)

7.3.2 Flash-beep task – accuracy

Accuracy scores were analysed in order to determine whether participants could accurately count the visual flash stimuli either in the absence of an auditory stimulus or with a congruent number of beeps, and then to determine the extent of fission and fusion illusions reported in trials where an illusion was expected.

For each of the six stimulus types, accuracy scores were calculated by dividing the number of correct responses by the total number of valid responses made, so that non-responses were not counted as incorrect. Non-responses were rare, with 95.6% valid responses for adults and 96.9% valid responses for children. For the non-illusion trials (1F0B, 1F1B, 2F0B, 2F2B), the accuracy scores reflected the degree to which participants were able to accurately count the visual flash stimuli with no beeps or with a congruent number of beeps. Conversely, for the illusion trials, low accuracy scores indicated the presence of illusory perception. In fission trials (1F2B), low accuracy indicated that more flashes were reported than were presented, and in fusion trials (2F1B), low accuracy indicated that less flashes were reported than were presented.

Figure 2 shows mean accuracy scores (with 95% confidence intervals) for each stimulus type. In general, all participants responded with a high level of accuracy for all non-illusion stimuli, indicating that the visual stimuli were not ambiguous and that the visual flashes could be counted relatively easily by both children and adults. This was the case both in trials with no auditory stimulus present (1F0B and 2F0B) and in those where the number of auditory and visual stimuli were equal (1F1B and 2F2B). However, accuracy was generally reduced for
both types of illusion trials (fission – 1F2B, and fusion – 2F1B) in both adults and children.

Figure 7-2. Mean accuracy measures and 95% confidence intervals (CI) for each stimulus type for adults and children.

In fission illusion trials (1F2B), participants often reported more flashes than were present; in fusion illusion trials (2F1B), participants often reported less flashes than were presented. While accuracy for adults and children dropped to a similar level in fusion trials, accuracy for children dropped far more than in adults for fission trials, suggesting that children may be more susceptible to the flash-beep illusion than adults, especially in fission trials. Indeed, children reported more flashes than were presented on almost all fission illusion trials.

The significance of these effects was assessed using a mixed analysis of variance (ANOVA) with a between-subjects factor of Group [adults, children], and within-subjects factors for Nflash [1 flash, 2 flashes], and Nbeep [0 beeps, 1 beep,
2 beeps]. There was a significant main effect of group, $F(1,50)=23.5, p<.001, \eta^2=.32$, with adults significantly more accurate than children overall. There were also significant nflash, $F(1,50)=11.8, p=.001, \eta^2=.20$, nbeep, $F(2,100)=55.2, p<.001, \eta^2=.52$, and Nflash x Nbeep x Group interactions, $F(2,98)=3.6, p=.03, \eta^2=.07$. Accuracy in the illusion trials compared to the corresponding non-illusion trials was examined by decomposing the three-way interaction using simple effects analysis (Howell, 2009). Pairwise comparisons with Bonferroni-corrected alpha levels are reported throughout. In one-flash trials, both adults, $F(2,49)=31.9, p<.001$, and children, $F(2,48)=98.5, p<.001$, had significantly lower accuracy scores on fission illusion trials compared with both non-illusory single-flash trials, although the extent of the drop in performance was significantly larger for children than adults, $F(1,48)=28.2, p<.001$.

In two-flash trials, there was again a significant drop in accuracy for both adults, $F(2,49)=12.5, p<.001$, and children, $F(2,49)=25.8, p<.001$, for the fusion illusion trials compared to both of the non-illusion two-flash trials. However, there was no significant difference in accuracy between adults and children in the fusion illusion trials.

Thus, the fission illusion, originally reported by Shams et al (2000), was found to be present in children, and to a significantly greater degree than in adults. The corresponding fusion illusion, first reported by Andersen et al (2004), was also found to be present in children, but not to a significantly greater degree than in adults.

### 7.3.3 Flash-beep – reaction times

In order to further explore the difference between fission and fusion trials in adults and children, reaction times for all six stimulus types were examined. Reaction times for correct and incorrect responses were collected. In the non-illusion trials, only correct responses were analysed. However, in the illusion trials, incorrect responses are also of interest, as they are linked to reports of illusory perception. The reaction time data were thus split into illusion trials (1F2B, 2F1B)
and non-illusion trials (1F0B, 1F1B, 2F0B, 2F2B). Mean reaction times (and 95% confidence intervals) for each stimulus type are shown in Figure 3. Consistent with the Pegboard results, children were consistently slower to respond than adults in the non-illusion trials (Figure 3A). However, as can be seen in Figure 3B, adults appeared to lose their expected reaction time advantage in the illusion trials, except in incorrectly-responded fusion-illusion trials (i.e., when they reported the illusion).

![Figure 7-3. Reaction times (ms) and 95% CIs for non-illusion (A) and illusion (B) trials. Reaction times for correct responses only are shown in A. In B, reaction times for correct (C) and incorrect (I) responses are shown separately.](image)

For non-illusion trials (Figure 3A), a Group [adults, children] x Nflash [1 flash, 2 flashes] x Nbeep [0 beeps, 1 beep, 2 beeps] mixed ANOVA showed that adults were faster overall than children, $F(1,50)=15.5$, $p<.001$, $\eta^2=.24$. The only other significant effect was nbeep, $F(1,50)=21.0$, $p<.001$, $\eta^2=.30$, with significantly faster reaction times when a beep was present compared to no beep. No other main effects or interactions were significant, indicating that, for both adults and children, reaction times were fairly similar whether a single or double flash was presented. As a whole, however, participants were significantly faster when flashes were presented along with a congruent number of beeps.
In illusion trials (Figure 3B), a Group [adults, children] x Stimtype [fission, fusion] x Response [correct, incorrect] repeated-measures ANOVA showed no main effect of group, $F(1,43)=.04, p=.8$, indicating that children and adults showed similar reaction times in illusion trials overall. However, there was a significant Group x Stimtype x Response interaction, $F(1,43)=4.5, p=.04, \eta^2=.10$. Simple effects analysis revealed that while reaction times to incorrectly-responded stimuli for adults and children did not differ in the fission trials, adults were significantly faster, $F(1,43)=7.1, p=.01$, than children when they responded incorrectly to fusion stimuli.

By comparing reaction times across all stimuli in panels A and B in Figure 3, it is also clear that while children appear to respond with roughly similar overall reaction times in illusion and non-illusion trials, adults are generally slower in illusion trials compared to non-illusion trials. Reaction times for all illusion and non-illusion stimuli were averaged together, and a Group [adults, children] x Illusion [illusion, no-illusion] repeated-measures ANOVA revealed a significant Group x Illusion interaction, $F(1,50)=16.4, p<.01, \eta^2=.25$. Simple effects analysis indicated that while there were no differences in reaction times between illusion and no-illusion trials for children, adults were significantly slower in illusion trials than no-illusion trials, $F(1,50)=38.3, p<.01$.

### 7.3.4 Flash-beep – correlations

To determine if there was any relationship within each group between age and performance on the illusion task, Pearson correlations were calculated between age and the accuracy scores and reaction times for all stimulus types. No significant correlations were found between age and any of these variables, for both adults and children. A representative example is given in Figure 4 for accuracy scores in the two illusion stimuli. Although adults and children showed very different results, there was no relationship within each group between age and accuracy for either the fission or fusion illusion trials.
Figure 7-4. Scatter plots showing raw accuracy scores for adults and children to all stimuli. Illusion trials are shown on the far right (top and bottom rows).

In order to explore any relationship between accuracy, reaction times, and Purdue Pegboard scores, Pearson correlations were also calculated between accuracy scores for fission and fusion stimuli, the pegboard scores for the right hand, and reaction times for correct and incorrect responses for both illusion stimuli. Table 2 shows the matrix of correlations performed with accuracy scores. The Purdue Pegboard scores did not significantly correlate with any accuracy or reaction time variables in either adults or children, indicating that manual dexterity does not have a differential effect for different stimulus combinations in either group. The Pegboard correlations are not included in the table to conserve space. Note that the N values for correlations involving reaction times vary between conditions. This is a consequence of examining reaction times for both correct and incorrect responses – participants with 100% or 0% accuracy did not submit incorrect- or correct-response reactions times, respectively. In adults, accuracy for fission and fusion trials was significantly correlated, with those participants reporting the fusion illusion also reporting fission illusions, and vice versa. Children showed no such relationship. The only significant correlations between accuracy scores and reaction times were in adults – adult participants who had high
accuracy for both fission and fusion stimuli had slower reaction times for correct responses to fusion stimuli.

Table 7-2 Correlations Between Accuracy Scores for Fission and Fusion Stimuli, Purdue Pegboard Scores, and Reaction Times (RT) for Correctly and Incorrectly Responded Stimuli.

<table>
<thead>
<tr>
<th></th>
<th>Accuracy – Adults</th>
<th>Accuracy - Children</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Fission</td>
<td>Fusion</td>
</tr>
<tr>
<td>Fission Accuracy</td>
<td>r</td>
<td>.53*</td>
</tr>
<tr>
<td></td>
<td>N</td>
<td>22</td>
</tr>
<tr>
<td>Fission RT(correct)</td>
<td>r</td>
<td>-.15</td>
</tr>
<tr>
<td></td>
<td>N</td>
<td>21</td>
</tr>
<tr>
<td>Fission RT (incorrect)</td>
<td>r</td>
<td>.17</td>
</tr>
<tr>
<td></td>
<td>N</td>
<td>21</td>
</tr>
<tr>
<td>Fusion RT (correct)</td>
<td>r</td>
<td>-.45 *</td>
</tr>
<tr>
<td></td>
<td>N</td>
<td>19</td>
</tr>
<tr>
<td>Fusion RT(incorrect)</td>
<td>r</td>
<td>-.20</td>
</tr>
<tr>
<td></td>
<td>N</td>
<td>21</td>
</tr>
</tbody>
</table>

* p < .05 (2-tailed)
** p < .01 (2-tailed)

7.4 Discussion

In the current study, adults and children were asked to count how many flashes they saw during the presentation of an auditory-visual illusion that indexes the integration of auditory and visual information. One or two flashes were presented, along with zero, one, or two beeps. The main finding was that children, performing the same task as adults, reported significantly more fission illusions than adults, where the presentation of two beeps caused a single flash to be reported as two. Children also reported more fusion illusions than adults, where a single beep caused a double flash to be reported as one, but the difference was not significant. Despite children reporting significantly more fission illusions than adults, there was no correlation within either group between age and accuracy on either fission or fusion trials. These results show that the mechanisms that integrate auditory
with visual information, giving rise to the flash-beep illusion, do not follow a linear developmental trend with age in this group of normally developing children.

Both adults and children reported fission and fusion illusions on trials where either more or fewer beeps than flashes were presented. Adults reported illusions on approximately 50% of trials, a figure comparable with many other studies using the flash-beep illusion paradigm (Andersen et al., 2004, Shams et al., 2000, Shams et al., 2002, Courtney et al., 2007, Meylan and Murray, 2007, Innes-Brown and Crewther, 2009). However, children reported far more fission illusions than adults. In the flash-beep illusion, the accuracy of counting a quickly flashed visual stimulus decreases if an incongruent number of beeps are presented at the same time. There have been many behavioural and neuro-imaging studies suggesting that this effect occurs as a result of the integration of auditory cues with visual information at a relatively low level in the brain, at least at the level of the primary visual cortex (Watkins et al., 2007, Watkins et al., 2006, Shams et al., 2005a, Shams et al., 2001, Mishra et al., 2008, Mishra et al., 2007). There is also psychophysical evidence suggesting that the phenomenon is the result of an illusory visual percept, rather than the modification of response biases (McCormick and Mamassian, 2008). In the current study, we have shown that children appear to integrate auditory and visual information more frequently and less selectively in the flash-beep illusion task than adults, leading to more illusory perceptions when incongruent auditory-visual stimuli are presented.

To our knowledge, there is only one previous study examining the flash-beep illusion in children (Tremblay et al., 2007). In that study, children in three age ranges (5-9, 10-14, and 15-19) performed flash-beep illusion as well as McGurk illusion tasks. The authors found that the youngest group of children reported significantly fewer McGurk illusions than either of the older groups. In the flash-beep task, however, no differences were found between the groups, although all the age groups did show significant fission and fusion illusions. The authors suggested that different mechanisms underly the flash-beep illusion, which was already mature in their sample, compared with the mechanisms underlying the McGurk effect, which was still developing. In the present study, it was not
possible to divide the children into three equally-sized age groups in order to make a direct comparison; however, no correlation was found within the child group between age and either accuracy or reaction time in the flash-beep task, and examination of the raw data (see Figure 4) does not suggest any linear, bimodal, or otherwise non-linear relation between age and accuracy. The results are thus consistent with those found by Tremblay et al., and the accuracy scores for the children in their study correspond well to those found in the current study. The current results also support their conjecture that the mechanisms underlying the two illusions are likely different. However, as shown in the current study, a group of adults performing the same task reported far fewer fission illusions than children, suggesting a different interpretation: that the mechanisms underlying the illusion are still developing.

As the main measure of illusory perception was a reduction in the accuracy of counting visual flashes, it was important to ensure that the reduced accuracy shown by children in illusion trials was not simply reflective of overall reduced accuracy in perceiving single and double flashes. In addition to the illusion trials, accuracy scores were recorded for single and double flash trials with either no concurrent beeps, or with a congruent number of beeps. In these control trails, adults were generally more accurate than children. However, although the accuracy of adults was approximately 13% higher than children in the 1F0B control trials (where the difference was largest), adult accuracy was 34% higher in the 1F2B illusion trials. Thus, although adults were more accurate overall, the difference was far more pronounced in the illusion trials. To fully control for this effect, a future study could equalise the task difficulty by the use of degraded flash stimuli for the adults.

Nevertheless, children’s accuracy rates were still high in the unisensory control trials – over 75% for one-flash trials and over 80% for two-flash trials, indicating that children had little difficulty in responding correctly to the visual stimulus when it was presented alone. Similarly, congruent multisensory trials were presented, where the number of beeps was the same as the number of flashes. In these trials, both adults and children were more accurate still, with all scores above
80%. It is well known that humans can detect multisensory combinations of signals more quickly and accurately than either of the signals presented alone. This facilitation of accuracy and reaction time is called the redundant signal effect, and has been extensively investigated in adults (Gielen et al., 1983, Molholm et al., 2004, Hecht et al., 2008). These studies have generally used Miller’s inequality (Miller, 1982) to show that reaction times to multisensory stimuli are faster than can be predicted by “race” models (Raab, 1962), in which the faster sense simply initiates the motor response. Instead, they suggest co-activation models, where the signal from each sense is integrated at an early stage in order to produce a faster response. Barutchu, Crewther, & Crewther (2009) and Barutchu et al (2010) used this method to show that although reaction times of children were generally faster with multisensory compared to unisensory stimuli, children up to 12 years of age were not as fast as adults and did not show the reaction time advantage that would be expected under a co-activation model. As in the present study, the level of multisensory facilitation observed did not correlate with either age or motor reaction times. In a relatively early study, Hulme, Smart, Moran & Raine (1983) also found no relationship between visual-kinaesthetic integration and motor skill development. In the current study, both children and adults were significantly faster in response to audio-visual than purely visual stimuli. However, an auditory-only stimulus was not included. This fact, and the very different task demands, makes it difficult to compare the results directly with studies explicitly designed to test Miller’s inequality. In any case, the participants in the current study showed significant benefits both in terms of accuracy and reaction times in response to visual stimuli when congruent audio-visual stimuli were presented compared to visual alone. Children had lower accuracy scores than adults in the visual-alone and congruent multisensory trials, and as expected from the lower Pegboard scores, were significantly slower overall. However, both groups showed high levels of accuracy overall, were significantly faster in congruent multisensory trials, and had no problems distinguishing a single flash stimulus from a double flash stimulus. Taken together, these results further support the conjecture that MSI does not
follow a linear developmental trajectory with either age or the development of motor coordination.

As well as reporting fewer illusions overall, adults had significantly faster reaction times than children in the unisensory and congruent multisensory control trials. In illusion trials, however, the adults maintained higher accuracy levels than the children, but lost the reaction time advantage and responded more slowly than in the control trials. This slowing in the illusion trials was especially apparent for trials in which a correct response was made (i.e. the illusion was not reported). In adults, there was also a significant negative correlation between accuracy and reaction time, but only for fusion stimuli with a correct response. This may indicate that responding correctly in the presence of an incongruent number of beeps requires the suppression of an automatic integrative process, thereby slowing the final response. The ability to inhibit pre-planned responses has been shown to develop slowly, with improvements from the ages of 6 to 20 (Band et al., 2000), and these improvements have been linked to the development of prefrontal cortex (Tamm et al., 2002). In the current study, the automatic integration of auditory with visual information may have been unsuccessfully inhibited in children. In the fusion trials with an incorrect response, adults were also significantly faster than the children, again suggesting a failure in inhibition.

A further difference between the adults and children in the present study was in relation to the pattern of results for fission and fusion stimuli. Although children reported significantly more fission illusions than adults, the difference between groups was less clear and not statistically significant for fusion stimuli. In addition, there was a significant positive correlation between accuracy scores for each illusion trial type in adults, but not children – while adults who reported many fission illusions also reported many fusion illusions, there was no such relationship in children. One possibility for this difference is due to the reliability of each sense modality in the task of distinguishing between rapid events. In their original studies, Shams et al. (2000, Shams et al., 2002) only reported fission illusions – they found that an illusion occurred only when the number of beeps exceeded the number of flashes and not vice versa. The well-established “modality
appropriateness” hypothesis (Welch and Warren, 1980) thus could not explain their results. According to this theory, the number of beeps should always dominate over the number of flashes, as audition provides more accurate temporal information than vision. However, this did not appear to be the case in Shams et al. (2000, Shams et al., 2002), and the authors proposed instead the “discontinuity hypothesis” – that the discontinuous stimulus in one modality alters the percept of the continuous stimulus in the other modality. The present results indicate that the auditory stimulus, whether it was continuous or discontinuous, dominated, or at least altered, the visual percept. Hence, the results are more supportive of the modality appropriateness hypothesis.

In the current study, the effect of the discontinuous stimulus (the double beep) was stronger in children than adults. These results are in agreement with some early data indicating that infants show auditory dominance when presented with an audio-visual checkerboard/beep stimulus. Infants are more responsive to changes in the rate of the beep stimulus than the spatially static flashing checkerboard (Lewkowicz, 1988a, Lewkowicz, 1988b) When the checkerboard is moving, however, the stimulus is spatially-dynamic, audition is no longer the more appropriate sense, and the responses are no longer dominated by the auditory stimulus (Lewkowicz, 1992). However, as Andersen et al. (2004) have also found, discontinuous auditory stimuli (i.e. the double beep in fission illusion trials) had a stronger effect on the final visual percept than continuous auditory stimuli (the single beep in fusion illusion trials), and Andersen et al. point out that it is likely that modality appropriateness and stimulus discontinuity are both factors that combine to influence the dominance of each modality.

The results from the present study are generally in agreement with previous research showing that the ability to appropriately integrate information from multiple senses is not fully developed until late in childhood. This finding is broadly consistent with neurophysiological studies in cat (Wallace et al., 2006) and monkey (for review see Stein et al., 2009b) which have characterised the development of multisensory neurons in the midbrain (superior colliculus - SC) and cortex (anterior ectosylvian sulcus – AES). These studies have found that
although multisensory neurons in the SC exist at birth in monkeys, or soon after in the cat, they have very large receptive fields that narrow with age, and although they respond to multiple sensory inputs, many do not have the ability to integrate these inputs (Wallace and Stein, 1997). As these receptive fields narrow with post-natal development, it is possible that integrative abilities become more selective. The ability of SC neurons to integrate inputs from multiple senses is slow to develop, and is guided by cortical input from the AES, an area of cortex which is slower to develop than the SC (Jiang and Stein, 2003, Stein et al., 2002). This AES-SC circuit is thought to be the mechanism by which experience and learning adapts multisensory responses in the SC to environmental pressures (Stein et al., 2009b). These findings fit with the idea that while some multisensory abilities, such as intensity matching (Lewkowicz and Turkewitz, 1980), audio-visual rhythm detection (Bahrick and Lickliter, 2000, Bahrick and Lickliter, 2004) and the detection of temporal equivalence in the form of speech onsets and offsets (Lewkowicz, 2010; Lewkowicz et al., 2010) can be present from birth, while other abilities, particularly those that require the integration of multiple sensory inputs, can require a prolonged developmental period. The stream-bounce illusion, for example, does not emerge in infants until six months of age (Scheier et al., 2003a), and spatial orienting responses do not show multisensory facilitation until 8 months of age (Neil et al., 2006). Facilitation of size and orientation judgements by visio-haptic stimuli occurs even later, at between 8-10 years of age (Gori et al., 2008), and children as old as 11 years do not show mature facilitation of reaction times to auditory-visual stimuli (Barutchu et al., 2010). The current study extends this time-line by showing that a group of children aged 8-17 reported significantly more fission illusions in the flash-beep illusion task.

Finally, it is important to note some major differences of approach between studies looking at multisensory facilitation and the current study. Overall, these studies have suggested that the ability to gain behavioural advantage from congruent multisensory stimuli does not mature until at least late childhood. The current study, in contrast, was focussed on incongruent multisensory trials, that, due to the integration of information from each sense, gave rise to perceptual illusions.
Hence, a reduction in accuracy was the main index of multisensory integration. Although children in the current study did not show mature performance in the multisensory illusion task, the results suggested that this was due to more, but less selective, integration of auditory with visual information, resulting in more illusory percepts. There are a multitude of possible neural mechanisms underlying the ability to integrate sensory information (Cappe et al., 2009b), and these mechanisms are likely to mature at different rates. In the context of the flash-beep illusion task, maturation of the sensory and multisensory pathways appears to result in the ability to more selectively integrate auditory information with visual information, resulting in less perceptual errors as development progresses. In the non-illusory control trials, it was also shown that both adults and children showed the expected facilitation of accuracy and reaction times scores to congruent multisensory compared to visual-alone stimuli.

### 7.5 Conclusion

This study has shown that the ability to selectively integrate auditory and visual information, as indexed by an auditory-visual illusion, is not mature in a group of 8-17 year-old children. In the flash-beep illusion task, children reported significantly more fission illusions than adults. Reports of the illusion were not correlated with age in either children or adults, suggesting a shift in development of this process at an age beyond the range included in this study. To further understand how these abilities develop, it may be necessary to test a variety of multisensory abilities in both adults and children over a wider range of ages. The current study is supportive of recent research suggesting that although multisensory abilities such as matching and comparison can develop very early in infancy, the ability to appropriately integrate sensory information in order to facilitate or alter sensory experience may require a much more extended period of development.
8 Discussion
8.1 Introduction

Many if not most stimuli in everyday life are multi-modal in nature. Most environmental events generate multiple congruent streams of information that our bodies can detect using more than a single sense modality, unless they are very far away, or unless we are deliberately insulated from specific modalities (by a window for example). In some situations these streams may be degraded or provide incomplete information by themselves. It is thus not surprising that we have evolved to take advantage of the redundancy inherent in these correlated streams in order to build up the most accurate possible interpretation of the available sense data. However, until fairly recently, the processing of sensory information in the brain was assumed to follow separate parallel streams, being progressively elaborated in these separate streams before being finally united in high level association cortices, such as those in the frontal or parietal lobes. As more sophisticated theories and analysis techniques have emerged, this view has given way to a new understanding that information is also transferred between sensory pathways at very early stages, at the level of the primary sensory areas and perhaps sub-cortically.

As was reviewed in Chapter 2, researchers have described facilitation of reaction times and discrimination thresholds for a variety of types of multi-modal stimuli. However, when the individual uni-modal parts of a multi-modal stimulus are in-congruent, perceptual illusions and alterations can occur. Situations such as these can arise in nature, for instance when an event is far enough away that the auditory and visual energy arrive at significantly different times, or in a laboratory, when an experimenter deliberately arranges incongruence between the uni-modal parts of a multi-modal stimulus. It is important to keep in mind that although the perception of multi-modal illusions is undoubtedly a result of multi-sensory processes in the brain, that these processes are actually erroneous, providing a perception of events that is not veridical (see Chapter 3, Section 3.1 for further discussion of this point). Nevertheless, the presence of multi-modal illusory percepts does provide strong evidence of underlying multi-sensory processes.
The three experiments presented in this thesis used the flash-beep illusion (Shams et al., 2000) as an index of multi-sensory processes in the brain – firstly to show that the illusion was robust to significant spatial separation of the auditory and visual parts of the stimulus, and secondly to show that congruent multi-modal stimuli initiated a sequence of long-lasting effects on subsequent uni-sensory processing that were localised to the parietal and occipital lobes. Lastly, it was shown that a group of children from 8-14 years of age reported significantly more illusions than adults, adding further weight to the growing body of evidence suggesting that many multi-sensory processes are slow to develop.

This discussion will draw together the essential findings from the experiments reported in Chapters 5, 6 and 7, and present them in the context of three possible brain network models explaining the results. Each experimental chapter has its own discussion section placing the results in the context of the existing literature, so only the key results of theoretical relevance to the proposed models will be repeated and elaborated upon here.

### 8.2 Models of multisensory processes

In Chapter 2, three models of multisensory processes were put forward based on experimental results and a review of the timing and location of multi-sensory effects found in similar experiments. Although more models have been suggested, the rest of this discussion will focus on these three possibilities. The models are summarised in Figure 8-1 below. In the first ‘traditional hierarchical’ model (Figure 8-1, panel A), sensory information is processed in separate, parallel pathways. Within each pathway, sensory information ascends from the sense organs, and progresses through sub-cortical areas to primary sensory areas. At each stage the information is further refined and elaborated, until it is finally brought together in parietal and frontal ‘association areas.’ Only after this point can feedback from one modality affect processing in the other modality. This has been the dominant model for understanding sensory processing until recently. In the ‘strict cortico-cortical’ model (Figure 8-1, panel B), only direct connections
between primary sensory areas are present. Figure 8-1 panel C shows a more likely ‘richly connected’ scenario, where existing hierarchical models are combined with more rapid, direct connections between primary sensory areas.

Figure 8-1. A) Diagram showing the location of the primary and secondary auditory (A1, A2) and visual (V1, V2) cortex in the human brain. V1 and V2 are located bilaterally on the occipital poles. A1 and A2 are only partly shown on this view, as they are located partly within the superior temporal
sulcus, on the superior surface. B) ‘traditional hierarchical’ view of the integration of auditory and visual information, where all sensory information is progressively elaborated throughout uni-sensory pathways before being integrated in higher ‘association cortex.’ C) Strictly cortico-cortical connections only D) Emerging view where, in addition to the flow of information to the higher areas of cortex (grey lines), there are also low-level connections directly between primary sensory cortices. The thick line indicates that evidence has been found for more numerous connections between the primary sensory areas than between secondary areas.

Reciprocal connections between A1-A2 and V1-V2 are presumed to exist in both models but are excluded for clarity. * A1 -> V1: (Rockland and Ojima, 2003, Falchier et al., 2002, Clavagnier et al., 2004). ** A1-V2 (Rockland and Ojima, 2003) *** V2-A1: (Budinger et al., 2006).

The flash beep phenomenon is a multi-modal illusion where auditory stimuli have the capacity to alter or influence visual perception. In terms of the models above, it is apparent that there are three main alternative routes by which the beep stimulus could cause the perception of an extra flash:

A) Auditory information from A1 is transferred to V1 by way of intermediate poly-sensory cortical areas;

B) Auditory information from A1 is transferred to V1 via direct cortico-cortical connections only;

C) A1 may communicate directly with V1 as in model 2), with additional influences from multi-sensory association areas in the parietal lobes.

It should be noted that there has also been support for a view where multi-sensory processes occur between sub-cortical structures, implicating thalamic regions such as the pulvinar (Hackett et al., 2007, Cappe et al., 2009a). However, the behavioural and EEG studies in this thesis were not equipped to rule in or out the involvement of sub-cortical structures, so this possibility will not be discussed
further. As reviewed in Chapter 2, the anatomic feasibility of the direct cortico-cortical connections involved in alternatives B) and C) is well-supported by tracer studies in non-human primates (Clavagnier et al., 2004, Falchier et al., 2002, Smiley and Falchier, 2009, Rockland and Ojima, 2003, Budinger et al., 2006) as well as more recent fMRI studies in humans (Eckert et al., 2008, Beer et al., 2011).

A timeline-based framework was introduced in Chapter 6 (Figure 6-13) in order to explain how long-lasting multi-sensory processes might affect subsequent processing of visual stimuli. In the framework proposed in Chapter 6, the various combinations of stimuli were separated into the first and second pairs of stimuli. For instance, the 2F1B stimulus studies were re-cast as an initial multi-modal (flash-beep) context stimulus closely followed by another uni-modal (flash) stimulus. By comparing responses to the second flash, depending on whether it was preceded by a uni-modal (flash) or multi-modal (flash-beep) context stimulus, it was possible to examine the lasting neural effects of the multi-modal context on subsequent uni-modal processing. In Chapter 6 illusion stimuli were not studied directly. However, it is possible to use the framework proposed in Chapter 6 in order to understand possible models for the generation of the illusory flash in the flash beep illusion. All that is necessary is to substitute the second flash with a hypothetical beep stimulus. In both cases the initial multi-modal flash/beep context stimulus is identical, so the pattern of neural responses initiated by it should also be the same. Thus, the framework can be used to predict the lasting effect of the multi-modal context stimulus on subsequent visual processing, with the addition of a second beep stimulus at 67 ms.

Figure 8-2 shows hypothetical patterns of neural interactions related to the flash-beep illusion stimulus (1F2B, or a multi-modal flash/beep followed by a second beep) in the framework of the three possible models. Overlaid are the neurophysiological results from Chapter 6 of this thesis (blue rectangles). Figure 8-2 is arranged in a similar manner to Figure 6-13 in Chapter 6. In panels A-C of Figure 8-2, hypothetical neural responses to the 1F2B fission illusion stimulus are shown. The initial flash/beep multi-modal context stimulus is at 0 ms, followed 67 ms later by a second beep (indicated by the grey vertical line). The initial neural
responses to the combined flash/beep stimulus, as well as to the second beep stimulus are shown in the primary auditory (A1) and visual (V1) cortices using green circles (auditory, points 1 & 3) and red circles (visual, point 2). As with Figure 6-13, the onset times for the first auditory cortex activity were taken from Lakatos et al (2005), and the for the visual cortex from Chen & Yeh (2009). The monkey timings from these studies were converted to human-equivalent times using the ‘3/5ths’ rule (Musacchia and Schroeder, 2009, Schroeder et al., 2004). It is acknowledged that this model is hypothetical, and all aspects are not directly tested by the experiments in this thesis. Rather, the model is an attempt to interpret the results, using a framework developed by examination of other empirical results from the literature.

As has been discussed in Chapter 6, Molholm et al (2006) have found that the activity related to a combined flash/beep stimulus in primary sensory areas is followed approximately 20 ms later by activity in the parietal lobes in each case (red and green inverted triangles, points 5 & 6). As the initial auditory and visual components of the stimulus were presented in a spatially and temporally congruent manner, there are also non-linear integrative responses in the same area starting from approximately 120-160 ms (overlayed red and green inverted triangles, point 7). The approximate time of perceiving the illusory second flash is indicated by the series of red circles (Figure 8-2, point 4). The timing of the illusory flash has been estimated by placing it at the time at which activity from a real flash at the same time as the second beep would have arrived at V1, using the same timing assumptions as above.
Figure 8-2. Three possible models of the generation of the illusory flash in the flash-beep illusion paradigm. Green and red circles represent the first onset of activation in the auditory (A1) and visual (V1) cortices by auditory (green circle, points 1 & 3) and visual (red circle, point 2) stimuli. Green and red inverted triangles show projected activity in the parietal lobe (PL) associated with the uni-modal stimuli. Overlayed green and red triangles show integrative response in the PL. Panel A: ‘traditional hierarchical’ model, where information from A1 and V1 is only combined in high-level parietal association areas. Panel B: ‘cortico-cortical only’ model, where information is combined directly between A1 and V1, with no involvement of high-level cortical areas. Panel C: ‘richly-connected’ model, where both cortico-cortical connections as well as feedback from high-level association
areas are involved. References: 1, 3: first onset of activity to broadband noise stimulus in macaque A1 (Lakatos et al., 2005). 2: response to light flash in macaque V1 (Chen et al., 2007). 5, 6, 7: first onset time of activity to auditory (5), visual (6), and integrative response to audio-visual stimulus (7) in human parietal lobe (Molholm et al., 2006).

While the timing and location of these hypothetical neural activities stays constant, the three models are differentiated by the possible flow of information between primary auditory and visual cortex as outlined above.

In the traditional hierarchical model (Figure 8-2, panel A), information feeds forwards from the primary sensory areas to the parietal association areas, before finally feeding back to the primary sensory areas. As can be seen from Figure 8-2, the timing of activations in A1 and the parietal lobes makes it difficult for this model to account for the flash-beep illusion. Activation related to the second flash at 67 ms occurs in A1 at approximately 80 ms. This is followed by activation in the parietal lobes around 30 ms later, from approximately 110 ms. For the illusory flash to occur at its predicted time (approximately 120-140 ms) would require almost instantaneous integrating responses in the parietal lobe and equally rapid communication to V1. In order to more fully evaluate the likelihood of this model, further data would be required: a more accurate estimation of the time at which the illusory flash is perceived, and a specific measure of the activity in V1 related to the illusory flash.

In the strict cortico-cortical connection model (Figure 8-2, panel B), only the ascending connections from primary sensory areas to association areas are present – there is no feedback from the association areas back to primary sensory areas. Instead, direct cortico-cortical connections feed information from primary auditory to primary visual cortex. Without top-down control from higher cortical areas, this arrangement may result in obligatory activation of visual areas for every beep stimulus.

In the richly-connected model (Figure 8-2, panel C), both feedback connections from association areas as well as direct cortico-cortical connections
are active. It is hypothesised that in this third model, feedback from parietal association areas, initially generated by the combined flash/beep stimulus at 0 ms, has the effect of priming V1. When V1 subsequently receives further input from A1 after the presentation of the second beep, this priming is enough to trigger a visual perceptual event (the illusory flash). Under this model, V1 requires both feedback modulation from association areas in the parietal cortex as well as activation by cortico-cortical connections from auditory cortex in order to generate the illusory extra flash. Although this hypothesis cannot be unambiguously proven or rejected by the data presented in this thesis, the results from this thesis and the other work as summarised below are generally supportive.

8.3 Spatial properties of the illusion related to neural spatial properties

The experiments reported in Chapter 5 showed that spatial separation of the auditory and visual components of the flash-beep stimuli by 20° of visual arc did not affect perception of the illusion. It should be acknowledged that the relatively small numbers (N = 9) in this study make definitive conclusions based on this negative finding difficult. Nevertheless, it was argued that this result implicated the involvement of cortical rather than sub-cortical brain regions, as while multisensory processes in sub-cortical areas such as the SC generally require the stimuli to be spatially and temporally congruent, the connections between A1 and V1 have been shown to less spatially selective.

Retrograde tracing studies in primates have found that the inputs from the core and parabelt auditory cortex to primary visual cortex provide their strongest connections to areas of primary visual cortex that subserve the peripheral visual field – from 10-20° eccentricity (Falchier et al., 2002). The flash-beep illusion is strongest when the flashes are presented in the periphery, and Shams et al (Shams et al., 2001) have shown that illusion-related modulations of the flash VEP occur only when the flashes are presented in the visual periphery. An eccentricity of 5-20° has been used in most studies of the illusion (see Table 3-1 in Chapter 3),
therefore most studies have presented stimuli within the range where connections from primary auditory to primary visual cortex subserve the visual periphery (Falchier et al., 2002). Although the minimum resolvable angle for sound localisation is only a few degrees, spatial receptive fields of individual neurons in the auditory cortex are large, generally occupying a quadrant or more of acoustic space with stimuli 10-30 dB above neural thresholds (Brugge et al., 2001). With higher sound levels, the receptive fields broaden still further (Mrsic-Flogel et al., 2005), despite higher-intensity sounds being easier to localise (Altshuler and Comalli, 1975). The experiment in Chapter 5 presented both auditory and visual stimuli on the left and right side in the visual periphery, but in some conditions the sides were swapped, such that the auditory stimulus was presented on the left and the visual stimulus on the right. In these spatially incongruent conditions, there was no effect on accuracy scores for any of the stimuli, indicating that perception of the illusion did not change despite the auditory stimulus being presented from a location 20° away from the visual stimulus. The results from Chapter 5 generally fit with the view that in order for the auditory stimulus to affect visual perception, integration of the auditory information would have to occur rapidly, perhaps via direct cortical pathways, with wide auditory receptive fields that are not sensitive to spatial information.

Overall, the results from Chapter 5 provided evidence that direct cortico-cortical connections from A1 to V1 may be involved in the flash-beep illusion. In terms of the models presented above, the results from Chapter 5 are generally supportive of models B or C – the two models that involve direct cortico-cortical connections. Model B however, is unlikely: without the modulatory effect of higher cortical areas, this arrangement would result in obligatory flash illusions for every beep stimulus, no matter what the previous context.

8.4 Involvement of feedback from parietal regions

The experiment described in Chapter 6 found that the electric field strength related to a flash stimulus was stronger when it was preceded by a multi-modal
flash/beep stimulus, compared to when it was preceded by another uni-modal flash stimulus. This difference was found to be significant in two distinct timeframes – an early timeframe, from 130-160 ms, and a late timeframe, from 300-320 ms. Source localisation analysis found that the increased activity in the early interval was localised to a diffused area centred on the inferior and superior parietal lobes, whereas the later difference was associated with stronger activity in a highly focussed area centred on primary and secondary visual cortex, in the occipital lobes. The timing and localisation of the effects suggested that the modulations of visual processing that were found may have been examples of feedback from parietal areas initially evoked by the multi-modal context stimulus.

The experiment in Chapter 6 emphasised the importance of the initial ‘context’ stimulus. When the context stimulus was multi-modal, subsequent neural responses to a visual stimulus were stronger than when the context stimulus was uni-modal. By re-conceptualising the stimuli in this way, a timeline was developed of the activations of key areas induced by the initial context stimulus (Figure 6-13). The activations of primary auditory and visual areas as well as association areas in the parietal cortex were based on existing literature as outlined in Section 6.3.3, and provided a framework for the understanding of the results. The early increase in activation found in the parietal lobe (Figure 8-2, point 8) was analogous to the activation of parietal association areas found at the same time and in the same location as in the cortical surface recordings of Molhom et al (2006). This timing and location is also consistent with MEG multi-sensory interactions found during the flash-beep illusion task (Shams et al., 2005a). Chapter 6 did not investigate the 1F2B illusion stimulus directly. However, the multi-modal context stimulus which initiated the changes in subsequent visual processing found in Chapter 6 is the same as is present in the first half of the 1F2B stimulus.

While Chapter 5 emphasised the possible involvement of direct, cortico-cortical connections between primary auditory and visual areas, Chapter 6 suggested that modulation of primary visual areas by parietal association areas may also be involved. The results from Chapter 6 thus support models A or C as described above – the two models that include inputs from parietal areas. When
the results are considered in the light of other research using the flash-beep illusion paradigm (as reviewed in the discussion of Chapter 6), model C is more likely – where parietal modulation ‘primes’ visual areas, just in time for the direct information from primary auditory areas to trigger the illusion.

8.5 Developmental aspects

Chapter 7 presented findings from translational research using the flash-beep illusion to index the development of multi-sensory processes in school-aged children. The main finding was that when performing the same flash-beep illusion task as adults, children reported significantly more fission illusions than adults. On average, children also reported more fusion illusions than adults, but this difference was not significant.

The results from Chapter 7 supported the hypothesis that the ability to appropriately integrate auditory and visual information is not fully developed until late in childhood. This finding is consistent with neurophysiological studies in cat (Wallace et al., 2006) and monkey (for review see Stein et al., 2009b) which have characterised the development of multisensory neurons in the midbrain (SC) and cortex (AES). Although the primate homologue of the AES is unknown, it has been linked with the posterior parietal cortex (Stein and Stanford, 2008). These studies have found that although multisensory neurons in the SC exist at birth in monkeys, or soon after in the cat, they have very large receptive fields that narrow with age, and although they respond to multiple sensory inputs, many do not have the ability to integrate these inputs (Wallace and Stein, 1997). As the receptive fields narrow with post-natal development, it is likely that integrative abilities become more selective. The ability of SC neurons to integrate inputs from multiple senses is slow to develop, and is guided by cortical input from the AES, an area of cortex which is slower to develop than the SC (Jiang and Stein, 2003, Stein et al., 2002). This AES-SC circuit is thought to be the mechanism by which experience and learning adapts multisensory responses in the SC to environmental pressures (Stein et al., 2009b). These findings fit with the idea that while some multisensory
abilities, such as intensity matching (Lewkowicz and Turkewitz, 1980), audio-visual rhythm detection (Bahrick and Lickliter, 2000, Bahrick and Lickliter, 2004) and the detection of temporal equivalence in the form of speech onsets and offsets (Lewkowicz et al., 2010, Lewkowicz, 2010) can be present from birth, other abilities, particularly those that require the integration of multiple sensory inputs, can require a prolonged developmental period.

The finding that adult-like performance in the flash-beep illusion task is slow to develop highlights the possible importance of cortical structures to the perception of the illusion (and guidance of multi-sensory processes more generally), as these structures are similarly slow to develop. The fact that children saw more illusions than adults suggests that without experience-guided control of higher-level cortical areas, the neural systems responsible for perception of the illusory flash tend to over-integrate. The results from Chapter 7 provide further support for Model C as described above, where modulation from cortical areas is critical for the illusion. However, the results from Chapter 7 provide a potentially new interpretation – as children saw more illusions than adults, these findings suggest either that the slow development of cortical areas caused the priming effect to be greater in children, or that the modulations from parietal areas may serve to inhibit automatic integration. Further research using brain imaging techniques in children across a wide range of ages will be required to disentangle these possibilities.

### 8.6 General conclusion

Overall, the experiments reported in this thesis highlight the involvement of high-level cortical association areas as well as direct cortico-cortical connections between A1 and V1 in the generation of the flash-beep illusion. It is proposed that both are required in order to generate the perception of an illusory flash: the preceding multi-modal flash/beep stimulus initiates a sequence of activity in the primary sensory areas leading to feedback from parietal areas priming V1. Direct,
spatially insensitive connections from A1 to V1 in the visual periphery may allow activity in A1 generated by the second beep to trigger a visual percept.

Neurons in auditory cortex have large spatial receptive fields, and the connections found by retrograde tracer studies in non-human primates show that A1-V1 connections project to the visual periphery. Relating the lack of spatial specificity found in Chapter 5 to the existing literature showing that the illusion only occurs when the visual stimuli are presented in the visual periphery suggests the involvement of connections between primary cortical areas in the generation of the illusion.

Chapter 6 used EEG source localisation techniques to show that an initial combined flash/beep stimulus initiated a series of multi-sensory modulations of subsequent uni-sensory processing of a flash stimulus. These effects were found in an early timeframe in the parietal lobes, from 130-160 ms, as well as a late timeframe in the occipital lobes, from 300-320 ms. The results were presented in a timeline-based framework which highlighted the importance of top-down feedback from parietal association cortex in addition to direct connections between primary sensory areas.

Finally, it was found that children report significantly more flash-beep fission illusions than adults. The results support a growing literature showing that multi-sensory processes develop slowly, and have implications for the design of teaching and learning environments. The fact that multi-sensory processes are slow to mature also highlights the importance of cortical, top-down control of lower level structures.

Further research in both the basic and translational arenas may benefit from the increased understanding that multi-sensory processes are generally the result of complex interactions between sub-cortical nuclei and primary sensory areas, guided by top-down control from high-level association regions, which are slow to develop during childhood.
9 References


Jurek, V., Tsuzuki, D. & Dan, I. 2007. 10/20, 10/10, and 10/5 systems revisited: Their validity as relative head-surface-based positioning systems. *Neuroimage*, 34, 1600-11.


10 Appendix 2 – Ethics approval letters, information sheets, and consent forms
10.1 Ethics approval

There were three main experiments run for this project.

1. A pilot psychophysics experiment run at the Brain Sciences Institute.
2. An extension of that experiment using spatialised stimuli, described in Chapter 5.
3. A further extension recording EEG during the presentation of the illusion (same ethics application).
4. A concurrently-run behavioural and EEG study with children, conducted at The Bionic Ear Institute, on the grounds of St Vincent’s Hospital in Melbourne. Only the behavioural data is presented in this thesis (Chapter 7), although an analysis of the EEG can be seen in the poster presentation attached in Appendix 2.

Ethics approval for was obtained from the Human Experimentation Ethics Committee at Swinburne University for (1) and (2), and from the Human Research Ethics Committee at the Royal Victorian Eye and Ear Hospital (the overseeing body for the Bionic Ear Institute), and the Catholic Education Office for (3). Copies of the relevant ethics committee approval letters, participant information sheets, and informed consent forms are provided below.
10.1.1 Advertisement

Sound-induced illusory flashing: a spatial exploration.

WHAT?

We’re conducting a study looking at how auditory and visual perceptions interact. Basically we will flash lights and simultaneously play beeps at you, and all you need to do is count the number of flashes you see. We need about 8 subjects, 4 males and 4 females. The session is broken into three parts, each of which will take about 10 minutes. No EEG, no drugs, no shocks, no blood taken. Your task, should you choose to accept it, is to press a number key on a keyboard at the end of each trial.

Inclusion criteria

1) Normal or corrected-to-normal colour vision and hearing
2) Right-handed
3) Nonsmoker
4) No history of epilepsy or any other major psychiatric disorder.

WHERE/WHEN?

The experiment will take place downstairs in David Crewther’s lab. You should allow an hour or so for the entire procedure. A mutually suitable session time will be arranged by email. No reimbursement is offered for this experiment.

CONTACT

Contact Hamish Innes-Brown by replying to this email, hinnesbrown@bsi.swin.edu.au

ph 9214 5543/8291

SMALL PRINT

This project aims to expand on an important recent finding in the science of auditory and visual perception. The first clear case where auditory information has radically affected unambiguous visual perception has recently been documented (Shams, Kamitani, and Shimojo, 2000; Shams, Kamitani and Shimojo, 2002). In these studies, varying the number of auditory beeps presented at the same time as visual flashes was found to vary the number of visual flashes perceived. Investigating the illusion can help us understand the way that auditory and visual information is combined in the brain.

After replicating the basic illusion the current project aims to expand on this result in two ways. Firstly we wish to examine the spatial properties of this cross-modal interaction – does varying the location of the cue flash with respect to the auditory beep affect the illusion? Shams et al (2000; 2002) used a spatially non-specific auditory stimulus. We plan to use auditory stimuli that are presented from either the same or opposite location as the visual flash. Following from this result our second intent is to use objects in space (rather than flashing disks on a computer display) that present both the auditory and visual stimuli. The convergence of these two experiments should provide interesting information regarding the brain systems that are involved with the illusion. If Shams et al (2000; 2002) are correct in saying that the illusion reflects a widespread property of polymodal mechanisms in the brain then the current results should go some way towards resolving whether these systems operate in a location- or object-specific manner.

10.1.2 Experiment 1 information sheet

THE AIM OF THIS STUDY

This project aims to expand on an important recent finding in the science of auditory and visual perception. The first clear case where auditory information has radically affected unambiguous visual perception has recently been documented (Shams, Kamitani and Shimojo, 2000; Shams, Kamitani and Shimojo, 2002). In these studies, varying the number of auditory beeps presented at the same time as visual flashes was found to vary the number of visual flashes perceived. Investigating the illusion can help us understand the way that auditory and visual information is combined in the brain.

After replicating the basic illusion the current project aims to expand on this result in two ways. Firstly we wish to examine the spatial properties of this cross-modal interaction – does varying the location of the one flash with respect to the auditory beep affect the illusion? Shams et al (2000, 2002) used a spatially non-specific auditory stimulus. We plan to use auditory stimuli that are presented from either the same or opposite location as the visual flash. Following from this result our second intent is to use objects in space (rather than flashing disks on a computer display) that present both the auditory and visual stimuli. The convergence of these two experiments should provide interesting information regarding the brain systems that are involved with the illusion. If Shams et al (2000, 2002) are correct in saying that the illusion reflects a widespread property of polysensory mechanisms in the brain then the current results should go some way towards resolving weather these systems operate in a location- or object-specific manner.


WHAT DOES THIS STUDY REQUIRE OF ME?

Participation in the study will take place at the Brain Sciences Institute (BSI). If you choose to participate in this study you will be asked to attend one session at the BSI to be arranged at your convenience. During this session you will complete three parts of the study. The entire testing session will take about one hour. In all sessions your task will be to count the number of flashes you see while some beeping noises are simultaneously presented. The only difference between the three parts of the experiment will be the way that the beeps and flashes are presented. In the first experiment flashes will be presented in the centre of a computer screen and the accompanying beeps from a speaker located under the monitor. In experiment two the flashes will occur in two locations on the screen, and beeps will come from speakers located to the left and right of the screen. In experiment three flashes and beeps will come from two combined beeping/flashing units located in front of you and slightly to the left and right.
DISCOMFORT AND POSSIBLE HAZARDS

You should not experience any discomfort during the session, nor should you be in any danger. Although the experiment is short, it is possible that you may experience a slight amount of boredom due to the repetition of beeps and flashes.

There is a very slight risk of epileptic seizure induction due to the flashing lights and abrupt sounds. If you have epilepsy or any history of seizure you will not be considered for participation.

Exclusion Criteria

We ask that individuals who suffer from epilepsy or who have a history of psychiatric illness do not participate in this study. You will also be briefly tested for normal hearing, and must have normal or corrected-to-normal vision.

Please do not hesitate to ask questions about the study, or any matter about it that concerns you, before, during and after your participation in the study. People you can ask include researcher Hamish Innes-Brown (ph. 0404 061 610) and supervisor Dr David Crewther (ph. 9214 5877). Before deciding on whether or not to participate in the study, you may wish to discuss the matter with a relative, friend or with your local doctor. You should feel free to do this.

Please note: You are free to withdraw from this study at any time, without consequence.

The results from this study may appear in publications or be provided to other researchers, but your identity will be kept strictly confidential at all times.

Should you have any questions regarding this study, please do not hesitate to ask the researchers whom you have met today. Alternatively, any further enquiries regarding the study can be directed to the Senior Investigator, Dr. David Crewther of the Brain Sciences Institute, on ph: 9214 5877. If you have any complaints or queries that have not been satisfactorily answered, please write to

Dr David Crewther
Acting Director of Brain Sciences Institute
Swinburne University of Technology
400 Burwood Road
HAWTHORN. VIC. 3122
Phone: (03) 92148273.

The Chair
Human Research Ethics Committee
Swinburne University of Technology
P O Box 218
HAWTHORN. VIC. 3122
Phone: (03) 9214 5223.
10.1.3 Experiment 1 consent form

INFORMED CONSENT FORM
SUBJECT CODE ______


Investigators: Dr David P Crewther, Hamish Innes-Brown
Brain Sciences Institute

Project Description:
We are conducting a research study to examine how vision interacts with sound.

As a participant, you will be asked to complete three short testing sessions where you task will be to count the number of flashes you see while some beeps are played at the same time.

The total time involved in the study will be about one hour.

Participation is voluntary and your initial agreement does not stop you from discontinuing the study at any time.

Your personal details will remain strictly confidential and your results will be identified by a code rather than a name. The results of the study may be published or shared with other researchers but your code, not your name will be used so in no way will the participants be identifiable.

Should you have any questions regarding the study titled “Sound-induced illusory flashing: a spatial exploration” you can direct them to Hamish Innes-Brown or to Dr David Crewther, the senior investigator.

Should you have a query that neither the researcher conducting the study or the senior researcher were able to satisfy, or any complaint about the way you were treated during the study, please write to:

The Chair
Human Experimentation Ethics Committee
Swinburne University of Technology
P O Box 218
HAWTHORN, VIC. 3122

Thank you very much for your participation
INFORMED CONSENT FORM

SUBJECT CODE ______


Investigators: Hamish Innes-Brown, BCogSci (hons)
Dr David Crewther, PhD.

I ___________________________ (the participant's name), have read and understood the information relating to this study. Any questions I have asked have been answered to my satisfaction.

I agree to participate in the research study titled Sound-induced illusory flashing: a spatial exploration. I realise that I may withdraw from the study at any time without explanation.

I agree that the research data collected for the study may be published or shared with other researchers on the condition that my name is not used and there is no other way in which I could be individually identified.

SURNAME: __________________________ OTHER NAMES: __________________________

SIGNED: __________________________ DATE: ____________

participant

NAME/S of PRINCIPAL INVESTIGATOR/S: __________________________

SIGNED __________________________ DATE: ____________

investigator(s)
10.1.4 Experiment 1 ethics approval

11 October 2002

A/Prof David Crewther
Brain Sciences Institute
Mail H99
Swinburne University of Technology
John St
Hawthorn, 3122

Dear A/Prof Crewther

Re: Sound-induced illusory flashing: a spatial exploration (2002/38)

The Joint School of Biophysical Sciences & Electrical Engineering and Brain Sciences Institute Human Research Ethics Sub-committee now has pleasure in advising you that your application for ethical clearance has been approved for 1 year.

Will you please inform the Committee of your progress at the end of 12 months or at the conclusion of the project. Continuation will be required for projects extending further than 12 months.

Yours faithfully

[Signature]

Prof Con Stough
Chair BSEE/BSI Human Research Ethics Sub-Committee

cc: Hamish Innes-Brown
10.1.5 Experiment 2 information sheet

THE AIM OF THIS STUDY

This project aims to expand on an important recent finding in the science of auditory and visual perception. The first clear case where auditory information has radically affected unambiguous visual perception has recently been documented (Shams, Kamitani and Shimojo, 2000; Shams, Kamitani and Shimojo, 2002). In these studies, varying the number of auditory beeps presented at the same time as visual flashes was found to vary the number of visual flashes perceived. Investigating the illusion can help us understand the way that auditory and visual information is combined in the brain.

In a previous project we have replicated the illusory effect. The next stage of this research aims to fully explore the electrophysiological correlates of the illusion – what happens to electrical activity in the brain while someone is experiencing the illusion. By looking at the differences in brain electrical activity, we can begin to understand the processes by which information is transferred between the senses.


WHAT DOES THIS STUDY REQUIRE OF ME?

Participation in the study will take place at the Brain Sciences Institute (BSI). If you choose to participate in this study you will be asked to attend one session at the BSI to be arranged at your convenience. The entire testing session will take about one hour. Your presence during the session will be to count the number of flashes you see while some beeping noises are simultaneously presented. An elastic cap with embedded electrodes will be placed on your head to record the electrical activity of your brain throughout the testing session.
DISCOMFORT AND POSSIBLE HAZARDS

You should not experience any discomfort during the session, nor should you be in any danger. Although the experiment is short, it is possible that you may experience a slight amount of boredom due to the repetition of beeps and flashes. The electrode cap should not be uncomfortable, and the skin will not be broken at any stage. An electrolyte gel is applied between the scalp and the electrode surface. This is basically a salt solution and will be washed out of your hair after the testing session.

There is a very slight risk of epileptic seizure induction due to the flashing lights and abrupt sounds. If you have epilepsy or any history of seizure you will not be considered for participation.

Exclusion Criteria

We ask that individuals who suffer from or who have a history of epilepsy or any other psychiatric illness do not participate in this study. You will also be briefly tested for normal hearing, and must have normal or corrected-to-normal vision. Individuals taking any form of psychoactive medication including nicotine will also not be considered for inclusion.

Please do not hesitate to ask questions about the study, or any matter about it that concerns you, before, during and after your participation in the study. People you can ask include researcher Hamish Innes-Brown (ph. 0404 061 610) and supervisor A/Prof David Crewher (ph. 9214 5877). Before deciding on whether or not to participate in the study, you may wish to discuss the matter with a relative, friend or with your local doctor. You should feel free to do this.

Please note: You are free to withdraw from this study at any time, without consequence.

The results from this study may appear in publications or be provided to other researchers, but your identity will be kept strictly confidential at all times.

Should you have any questions regarding this study, please do not hesitate to ask the researchers whom you have met today. Alternatively, any further enquires regarding the study can be directed to the Senior Investigator, A/Prof. David Crewher of the Brain Sciences Institute, on ph: 9214 5877. If you have any complaints or queries that have not been satisfactorily answered, please write to

A/Prof. David Crewher  The Chair
Director of Brain Sciences Institute  Human Research Ethics Committee
Swinburne University of Technology  Swinburne University of Technology
400 Burwood Road.  P O Box 218
HAWTHORN, VIC. 3122  HAWTHORN, VIC. 3122
Phone: (03) 92148273.  Phone: (03) 9214 5223.
Participant Info Sheet

Thanks for agreeing to be part of this study. This sheet will give you an overview of what’s required for the day.

On the day of testing you will be given two documents:

1. A “plain english statement” which briefly describes the project, goes over what’s required of you for the testing session, and says that you are here voluntarily and are free to withdraw your consent to participate at any time.

2. A “consent form” which you will need to sign to let us use the data collected during the session.

There are a few things that you can do to make the recording go more smoothly:

1. Have clean dry hair. Just like group-photo day in primary school! If you wash your hair the night before, and don’t use any styling products (even conditioner, if possible!), the electrode application procedure will go much, much faster. Applying the electrodes takes almost as much time as the testing procedure itself, and having clean dry hair can easily halve the application time. I’ll wash out the electrode gel afterwards, so bring some product with you if you’re going out into the wide world afterwards.

2. Don’t be too hung over or wired on caffeine. You don’t need to abstain, but I won’t be able to test anyone who has the shakes (caused either from the night before, or the jug of coffee you had 5 mins before testing).

3. If you have contact lenses and sometimes get dry eyes, either bring eye drops or a pair of glasses. Excessive blinking is bad. It’s also quite dry in the lab.

How to get here. The lab I’m using is on the 1st floor. When you get in I will hopefully be waiting for you. If I’m not there, use the phone in the foyer to ring 5656 or 5542, and I’ll come down and get you. If all else fails, my mobile is 0404 061 610. Here is a map:

Here I am!
BRAIN SCIENCES INSTITUTE
400 Burwood Road
Hawthorn
10.1.6 Experiment 2 consent form
Project title: Sound-induced illusory flashing: electrophysiology of the illusion.

Investigators: Hamish Innes-Brown, A/Prof. David P Crewther.
Brain Sciences Institute

Project Description:

We are conducting a research study to examine how vision interacts with sound.

As a participant, you will be asked to complete three short testing sessions where your task will be to count the number of flashes you see while some beeps are played at the same time. The electrical activity of your brain will be recorded while you complete the task.

The total time involved in the study will be about ninety minutes.

Participation is voluntary and your initial agreement does not stop you from discontinuing the study at any time.

Your personal details will remain strictly confidential and your results will be identified by a code rather than a name. The results of the study may be published or shared with other researchers but your code, not your name will be used so in no way will the participants be identifiable.

Should you have any questions regarding the study titled “Sound-induced illusory flashing: electrophysiology of the illusion” you can direct them to Hamish Innes-Brown or to Dr David Crewther.

Should you have a query that neither the researcher conducting the study or the senior researcher were able to satisfy, or any complaint about the way you were treated during the study, please write to:

The Chair
Human Experimentation Ethics Committee
Swinburne University of Technology
P O Box 218
HAWTHORN, VIC. 3122

Thank you very much for your participation
INFORMED CONSENT FORM

SUBJECT CODE

Project title: Sound-induced illusory flashing: electrophysiology of the illusion.

Investigators: Hamish Innes-Brown, BCogSci (hons)
A/Prof. David Crewther, PhD.

I, ___________________________________________ (the participant’ name), have read and understood the information relating to this study. Any questions I have asked have been answered to my satisfaction.

I agree to participate in the research study titled Sound-induced illusory flashing: phase two. I realise that I may withdraw from the study at any time without explanation.

I agree that the research data collected for the study may be published or shared with other researchers on the condition that my name is not used and there is no other way in which I could be individually identified.

To my knowledge I do not suffer from epilepsy nor have a history of seizure.

SURNAME: ___________________________ OTHER NAMES: ___________________________

SIGNED: ______________________ DATE: ____________

participant

NAME/S of PRINCIPAL INVESTIGATOR/S: ___________________________

SIGNED ______________________ DATE: ____________

investigator(s)
10.1.7 Experiment 2 ethics approval

15th April 2005

Hamish Innes-Brown
Brain Sciences Institute
Faculty of Life and Social Sciences

Dear Hamish,

Re: Sound-induced illusory flashing: phase two. (2005/2)

The Brain and Neurosciences Ethics Subcommittee of the Faculty of Life and Social Sciences has now considered your application for ethical clearance and takes pleasure in advising you that your application has been approved until 14th November 2005.

Will you please inform the Committee of your progress at the end of 12 months or at the conclusion of the project. An application for continuation will be required for projects extending further than 12 months.

Yours faithfully

[Signature]

Dr Con Stough
Chair Ethics Committee

cc: David Crewther
Rodney Croft
File

15/04/2005
10.1.8 Experiment 3 information and consent form

The Bionic Ear Institute
Royal Victorian Eye & Ear Hospital

PARTICIPANT INFORMATION AND CONSENT FORM

Version: 6
Dated: 10/11/08
Site: The Bionic Ear Institute,
6th Floor Daily Wing, St Vincent’s Hospital,
Victoria Parade, Fitzroy, VIC 3065.
Project Title: Multisensory processing in children and adults: effects of sensory and language disabilities.
Principal Researcher: Dr David B Grayden
Associate Researcher(s): Ms Ayala Barutchu, Mr Hamish Innes-Brown, Associate Professor Sheila Crowther, Dr Shani Dettman, Mr Mohit Shivasani, Ms Sarah Toohey, Ms Joanne Pifer, Ms Jaclyn Danaher, Ms Harriet Downing, Associate Professor Antonio Paolini.

Please make sure you have all 13 pages of this Participant Information and Consent Form.

1. Your Consent
   You are invited to take part in this research project. This Participant Information contains detailed information about the research project. Its purpose is to explain to you as openly and clearly as possible all the procedures involved in this project before you decide whether or not to take part in it.
   Please read this Participant Information carefully. Feel free to ask questions about any information in the document. You may also wish to discuss the project with a relative or friend or your local health worker. Feel free to do this.
   Once you understand what the project is about and if you agree to take part in it, you will be asked to sign the Consent Form. By signing the Consent Form, you indicate that you understand the information and that you give your consent to participate in the research project. You will be given a copy of the Participant Information and Consent Form to keep as a record.

2. Purpose and Background
   The purpose of this project is to learn how the brain combines sounds and images to obtain a coherent and unified perception of the world. We know that children and adults process information differently and we want to discover what brain changes are occurring as a child matures.
   There is some evidence that individuals with hearing impairment or a language-related learning difficulty process sounds and combine sounds and images differently to individuals without an impairment. We want to find out more about these differences and what relationship there is with other listening, language and thinking abilities. An ultimate goal of this research is to help individuals to learn to hear and speak better, based on a more complete understanding of the brain processes involved.
   A total of 200 people will participate in this project. You have been invited to participate because we want to compare adults and children, individuals with normal hearing and hearing loss and compare individuals with difficulty in language-related learning with those who do not.

3. Procedures
   Participation in this project will involve coming to our labs located at The Bionic Ear Institute on the 6th Floor Daily Wing, St Vincent’s Hospital, Victoria Parade, Fitzroy. You may be asked to attend up to 3 test sessions, depending on your age. Test sessions will generally last 1 hour. If you need to attend multiple sessions we will try to schedule them within a few weeks of each other. For children and teenagers we will try to schedule test sessions out of school hours.
   You will be asked to perform a brief screening test of hearing and vision. If you have a known hearing impairment you will be asked to perform only the vision screening test. Only individuals who pass the screening test will be asked to participate further in the research.
Following this you will be asked to:

- do some tests to assess how you process sounds and images. For example, you may be asked to indicate that you have heard or seen a sound or image or you may be asked to count the number of sounds or images. The researcher will carefully explain the task to you;
- wear an electrode cap which allows us to record your electrical brain activity while you hear sounds and see images;
- do some tests designed to assess thinking, reasoning, memory and speed of processing.

If you attend the Cochlear Implant Clinic at the Royal Victorian Eye and Ear Hospital and you give us your permission by signing the Consent Form, we will obtain from them details regarding your hearing impairment, cochlear implant, speech perception, language and abilities test results.

4. Possible Benefits

It is hoped that this research will ultimately help individuals with hearing impairment and language-related learning disabilities to hear and speak better. If you have such a condition you may eventually benefit from this research. However, we cannot guarantee or promise that you will receive any benefits from this project.

Your hearing and vision will be tested at the beginning of this research. If we find a deficit in your hearing or vision which you did not already know about we will refer you to an appropriate professional for further examination and advice.

5. Possible Risks

In order to obtain clear recordings of brain activity you will be asked to wash your hair prior to the research and not to use hair conditioner or hair products. The electrode cap used to record your electrical brain activity is quite tight and you may experience some discomfort as a consequence. All efforts will be made to minimise any discomfort and you can ask to remove it at any time. You can suspend or even end your participation in the research at any time.

You may unintentionally be exposed to uncomfortably loud sounds during the testing. If this occurs, testing will be stopped immediately and the volume modified. There may be additional unforeseen or unknown risks.

6. Privacy, Confidentiality and Disclosure of Information

Any information obtained in connection with this research project that can identify you will remain confidential and will only be used for the purpose of this research project. It will only be disclosed with your permission, except as required by law. If you give us your permission by signing the Consent Form, we plan to publish the results in academic journals. In any publication, information will be provided in such a way that you cannot be identified.

Information about you will be stored at the Bionic Ear Institute. It will be labelled in a way that does not identify you and will be accessed only by research staff. After 7 years, it will rendered unreadable.

7. Results of Project

It is usual for a number of years to pass before definitive results of this type of study are available. These are published in medical journals which are available to the public.

8. Further Information or Any Problems

If you require further information or if you have any problems concerning this project you can contact the principal researcher, Dr David Grayden or the senior research assistant, Ms Ayla Barutchu. The researchers responsible for this project are:

Dr David B Grayden
Ms Alya Barutchu
Mr Mohit Shivasani
Dr Shani Dettman
Assoc. Prof Sheila Crewther
Mr Hamish Innes-Brown
Ms Joanne Fifer
Ms Sarah Tooley
Ms Jaclyn Danaher
Ms Harriet Downing
Associate Professor Antonio Paolani

8344 5234
9288 3523
9288 3525
9929 8090
9479 2290
9288 3523
9288 3522
9288 3523
9288 3523
9288 3523
9288 3523
9288 3515
9. Other Issues

If you have any complaints about any aspect of the project, the way it is being conducted or any questions about your rights as a research participant, then you may contact:

Name: Kerryn Baker
Position: Secretary, Human Research and Ethics Committee
Telephone: (03) 9929 8525

You will need to tell Kerryn the name of one of the researchers given in section 8 above.

10. Participation is Voluntary

Participation in any research project is voluntary. If you do not wish to take part you are not obliged to. If you decide to take part and later change your mind, you are free to withdraw from the project at any stage. Your decision whether to take part or not to take part, or to take part and then withdraw, will not affect your routine treatment, your relationship with those treating you or your relationship with the Royal Victorian Eye and Ear Hospital.

Before you make your decision, a member of the research team will be available so that you can ask any questions you have about the research project. You can ask for any information you want. Sign the Consent Form only after you have had a chance to ask your questions and have received satisfactory answers.

If you decide to withdraw from this project, please notify a member of the research team before you withdraw. This notice will allow that person or the research supervisor to inform you if there are any health risks or special requirements linked to withdrawing.

11. Reimbursement for your costs

You will not be paid for your participation in this project. However, you will be reimbursed for travelling expenses and the cost of parking incurred as a result of participating in this research. In the case of car travel, reimbursement will be at the per km rate specified by The Bionic Ear Institute.

12. Ethical Guidelines

This project will be carried out according to the National Statement on Ethical Conduct in Research Involving Humans (June 1999) produced by the National Health and Medical Research Council of Australia. This statement has been developed to protect the interests of people who agree to participate in human research studies.

The ethical aspects of this research project have been approved by the Human Research Ethics Committee of the Royal Victorian Eye & Ear Hospital.

13. Injury

In the event that you suffer an injury as a result of participating in this research project, hospital care and treatment will be provided by the public health service at no extra cost to you.
Royal Victorian Eye & Ear Hospital

EXPERIMENTAL SUBJECT'S STATEMENT OF RIGHTS

The Royal Victorian Eye and Ear Hospital considers it important that you know:

Any patient who is asked to participate in a research study involving medical experiment, or who is requested to consent on behalf of another, has the right to:

1. Be informed of the nature and purpose of the experiment.
2. Be given an explanation of the procedures to be followed and any drugs used in the medical experiment.
3. Be given a description of discomforts and risks reasonably expected from the experiment, if applicable.
4. Be given an explanation of any benefits to the subject reasonably to be expected from the experiment, if applicable.
5. Be advised of appropriate, alternative procedures, drugs, or devices that might be advantageous to the subject, and their relative risks and benefits.
6. Be informed of the avenue of medical treatment, if any, available to the subject after the experiment if complications should arise.
7. Be given an opportunity to ask questions concerning the experiment or the procedures involved.
8. Know that consent to participate in the medical experiment may be withdrawn at any time, and that the subject may discontinue participation in the medical experiment without prejudice.
9. Be given a copy of the signed and dated written consent form when one is required.
10. Be given the opportunity to decide to consent or not to consent to a medical experiment without the intervention of any element of force, fraud, deceit, duress, coercion or undue influence.
The Bionic Ear Institute  
Royal Victorian Eye & Ear Hospital  

CONSENT FORM

Version: 6  
Dated: 10/11/08  
Site: The Bionic Ear Institute,  
6th Floor Daly Wing, St Vincent’s Hospital,  
Victoria Parade, Fitzroy, VIC 3065  
Project Title: Multisensory processing in children and adults:  
effects of sensory and language disabilities.

• I have read, and I understand the Participant Information version 6 dated 10/11/08.
• I freely agree to participate in this project according to the conditions in the Participant Information.
• I will be given a copy of the Participant Information and Consent Form to keep.
• The researcher has agreed not to reveal my identity and personal details if information about this project is published or presented in any public form.

Participant’s Name (printed) .................................................................
Signature .................................................................................................. Date

Name of Witness

to Participant’s Signature (printed) ............................................................
Signature .................................................................................................. Date

Declaration by researcher*: I have given a verbal explanation of the research project, its procedures and risks and I believe that the participant has understood that explanation.

Researcher’s Name (printed) .................................................................
Signature .................................................................................................. Date
The Bionic Ear Institute
Royal Victorian Eye & Ear Hospital

CONSENT FORM FOR MINORS

Version: 6
Dated: 10/11/08
Site: The Bionic Ear Institute,
6th Floor Daly Wing, St Vincent’s Hospital,
Victoria Parade, Fitzroy, VIC 3065

Project Title: Multisensory processing in children and adults: effects of sensory and language disabilities.

- I have read and I understand the Participant Information version 6 dated 10/11/08.
- I give my permission for ______________ to participate in this project according to the conditions in the Participant Information.
- I will be given a copy of Participant Information and Consent Form to keep.
- The researcher has agreed not to reveal the participant’s identity and personal details if information about this project is published or presented in any public form.

Participant’s Name (printed)

Name of Person giving Consent (printed)

Relationship to Participant

Signature Date

Name of Witness

to Parent/Court Appointed Guardian

Signature Date

Declaration by researcher*: I have given a verbal explanation of the research project, its procedures and risks and I believe that the participant’s parent/court appointed guardian has understood that explanation.

Researcher’s Name (printed)

Signature Date
10.1.9 Experiment 3 ethics approval

In reply please quote:

GE07/0009
1304
7 June 2007

Dr T Paolini
Auditory Clinical Neuroscience Unit
The Bionic Ear Institute
384–388 Albert Street
EAST MELBOURNE VIC 3002

Dear Dr Paolini

I am writing in response to your letter of 31 May 2007 in which you referred to your forthcoming research project titled *Multisensory processing in children and adults: Effects of sensory and language disabilities*. You have asked approval to approach Catholic schools in the Archdiocese of Melbourne, as you wish to recruit participants aged 6–16 years and their parents.

I am pleased to advise that your research proposal is approved in principle subject to the following standard conditions:

1. The decision as to whether or not research can proceed in a school rests with the school’s principal. So you will need to obtain approval directly from the principal of each school that you wish to involve.

2. You should provide each principal with an outline of your research proposal and indicate what will be asked of the school. A copy of this letter of approval, and a copy of notification of approval from the university’s Ethics Committee, should also be included.

3. A Working with Children (WWC) check – or an existing Criminal Record check – is necessary for all researchers visiting schools and must be shown to the principal before starting the research in each school. Details about applying for a WWC check may be obtained from the Department of Justice website <www.justice.vic.gov.au>.

4. No student is to participate in the research study unless s/he is willing to do so and informed consent is given in writing by a parent/guardian.

5. You should provide the names of schools which agree to participate in the research project to the Knowledge Management Unit of this Office.

6. Any substantial modifications to the research proposal, or additional research involving use of the data collected, will require a further research approval submission to this Office.
7. Data relating to individuals or schools are to remain confidential.

8. Since participating schools have an interest in research findings, you should consider ways in which the results of the study could be made available for the benefit of the school communities.

9. At the conclusion of the study, a copy or summary of the research findings should be forwarded to this Office. It would be appreciated if you could submit your report in an electronic format using the email address provided below.

I wish you well with your research study. If you have any queries concerning this matter, please contact Mr Mark McCarthy of this Office. The email address is <kmi@epo.melb.catholic.edu.au>.

Good wishes

Yours sincerely

[Signature]

Terri Hopkins
ACTING ASSISTANT DIRECTOR
POLICY AND GOVERNANCE
September 21, 2006

Dr Tony Paolini
The Bionic Ear Institute
384-388 Albert Street
East Melbourne Vic 3002

Dear Dr Paolini

Re: Human Research & Ethics Committee
  Project Number: 06/689H - Multisensory processing in children and adults:
  effects of sensory and language disabilities

I acknowledge receipt of Dr Shani Dettman from the RVEEH Cochlear Implant Clinic, joining your project.

Thank you for notifying the Committee. Your project may now proceed.

Yours sincerely,

Kerryn Baker
Secretary
Human Research & Ethics Committee
Telephone +61 3 9929 8525
Facsimile +61 3 9963 7203
kerryn.baker@eyeandear.org.au

cc HREC file
11 Appendix 2 – Publications.

Appendix 2 is unable to be reproduced online.
Please consult the print copy held in the Swinburne Library.