Putting it into Words: The Impact of Visual Impairment on Perception, Experience and Presence

By

Louise Fryer

Thesis submitted for the degree of Doctor of Philosophy

Department of Psychology

Goldsmiths College

University of London

2013

The work presented in this thesis is my own and any information or ideas sourced from the work of others are fully referenced within the text.

Signed: Date:

(Louise Fryer)

"Cognition is beautiful. It is beautiful to know" – John Hull (1935 -)

Acknowledgments

I would like to thank my supervisors, Dr. Jonathan Freeman and Prof. Linda Pring: Jonny for challenging me with wide-ranging discussions and brain-storming, Linda for her attention to detail and demand for clarity in research questions and design. Together they made the perfect team. Rob Davies gave invaluable assistance with the physiological measurements in the final study. Radio producer Amber Barnfather provided the stimuli for Study 3. I have also benefitted from David Hewlett's graphic design expertise. I am eternally grateful to Andrew Holland for first involving me in Audio Description (AD) and for our endless debates on the subject. For similar discussions I thank Roz Chalmers and all the describers I have worked with over many years. I thank, too, those academics in AVT who share my research interests, especially Pablo Romero Fresco. This research would not have been possible without all the sighted friends, neighbours and colleagues, who have given their time to take part. In particular, I thank all the blind and partially sighted participants who have given not only time but also invaluable insights. Special mentions go to Kirin Saeed for being a willing guinea pig at the design stage of several experiments; Mick Marr for being my companion at so many performances, described and not described, and John Thomas for helping me make the connection between AD and presence in the first place. Last but not least I thank my family: Steve, who made the haptic stimuli for Study 1, the haptic scale reader for Study 5 and assisted with data collection; Tom who edited the AV stimuli for Study 5; and Kez who, with the others, has had to endure ceaseless debates around the dinner table related to aspects of my research. That they have done so for 4 years while still retaining a vestige of interest is a great tribute to them and has been a huge support to me.

Abstract

The experience of being “present” in a mediated environment, such that it appears real, is known to be affected by deficits in perception yet little research has been devoted to disabled audiences. People with a visual impairment access audiovisual media by means of Audio Description, which gives visual information in verbal form. The AD user plays an active role, engaging their own perceptual processing systems and bringing real-world experiences to the mediated environment.

In exploring visual impairment and presence, this thesis concerns a question fundamental to psychology, whether propositional and experiential knowledge equate. It casts doubt on current models of sensory compensation in the blind and puts forward an alternative hypothesis of linguistic compensation. Qualitative evidence from Study 1 suggests that, in the absence of bimodal (audio-visual) cues, words can compensate for missing visual information. The role of vision in multisensory integration is explored experimentally in Studies 2 and 3. Crossmodal associations arising both from direct perception and imagery are shown to be altered by visual experience. Study 4 tests presence in an auditory environment. Non-verbal sound is shown to enhance presence in the sighted but not the blind. Both Studies 3 and 4 support neuroimaging evidence that words are processed differently in the absence of sight. Study 5, comparing mental spatial models, suggests this is explained by explicit verbal encoding by people with a visual impairment. Study 6 tests the effect of words on presence and emotion elicitation in an audiovisual environment. In the absence of coherent information from the dialogue, additional verbal information significantly improves understanding. Moreover, in certain circumstances, Audio Description significantly enhances presence and successfully elicits a target emotion. A model of Audio Description is presented. Implications are discussed for theoretical models of perceptual processing and presence in those with and without sight.

Parts of this thesis have appeared in the following publications:

Fryer, L. (2013). An ecological approach to audio description. *The Psychologist, 26* (6), 458 – 460.

Fryer, L. & Freeman, J. (2012). Presence in those with and without sight: implications for virtual reality and audio description. *Journal of CyberTherapy & Rehabilitation. 5*, 15-23.

Fryer, L. & Freeman, J. (2012). Cinematic language and the description of film: Keeping AD users in the frame. *Perspectives: Studies in Translatology*, DOI:10.1080/0907676X.2012.693108

Fryer, L., & Freeman, J. (2013, June). Visual impairment and presence: measuring the effect of audio description. In *Proceedings of the 2013 Inputs-Outputs Conference: An Interdisciplinary Conference on Engagement in HCI and Performance* (p. 4). ACM.

Fryer, L., & Freeman, J. (in press). Can you feel what I'm saying? The impact of verbal information on emotion elicitation and presence in people with a visual impairment. In A. Felnhofer and O. D. Kothgassner (Eds.) *Challenging Presence: Proceedings of the 15th International Conference on Presence*. Wien: facultas.wuv.

Fryer, L., Freeman, J., & Pring, L. (2013). What verbal orientation information do blind and partially sighted people need to find their way around? A study of everyday navigation strategies in people with impaired vision. *British Journal of Visual Impairment 31* (2), 123-138.

Fryer, L., Freeman, J., & Pring, L. (in press). Touching words is not enough: How visual experience influences haptic-auditory associations in the "Bouba-Kiki" effect. *Cognition*.

Fryer, L., Pring, L., & Freeman, J. (2013). Audio drama and the imagination: the influence of sound effects on presence in people with and without sight. *Journal of Media Psychology: Theories, Methods, and Applications 25* (2), 65-71.

Romero-Fresco, P., & Fryer, L. (2013). Could audio described films benefit from Audio Introductions? An audience response study. *Journal of Visual Impairment and Blindness 107 (4),* 287- 285.

Contents

Contents continued

Contents continued

Contents continued

Contents continued

Contents continued

Contents continued

Contents continued

List of Figures

List of Figures continued

List of Figures continued

List of Tables

List of Tables continued

Glossary of abbreviations

AD	Audio Description
ANOVA	Analysis of Variance
AV	Audiovisual
B	Blind
B&PS	Blind and Partially Sighted
BiP	Break in Presence
CAIS	Clarity of Auditory Imagery Scale
CAN	Central Autonomic Network
CB	Congenitally Blind
EB	Early Blind
ECG	Electrocardiography
EDA	Electrodermal Activity
EES	Emotion Elicitation Scale
ERK	Event-Related Knowledge
fMRI	Functional Magnetic Resonance Imaging
H	Hypothesis
HR	Heart Rate
HRV	Heart Rate Variability
HV	Human Voice
ITC-SOPI	ITC – Sense of Presence Inventory
LB	Late Blind
m	mean
md	mean difference

mdn	Median
MDS	Macular Disease Society
O&M	Orientation and Mobility
PERF	Primary Egocentric Reference Frame
PNS	Parasympathetic Nervous System
PS	Partially Sighted
RMSSD	Root Mean Square Deviations
RNIB	Royal National Institute of Blind People
rTMS	Repetitive Transcranial Magnetic Stimulation
S	Sighted
SD	Standard Deviation
SEM	Standard Error of the Mean
SFX	Sound Effects
SNS	Sympathetic Nervous System
SSM	Spatial Situation Model
TTS	Text-to-Speech
VE	Virtual Environment

Chapter 1 Introduction: Theories of Perception, Experience and Visual Processing

1.1 Background to thesis

Over the past thirty years, much research effort has been devoted to understanding the way people engage with "mediated environments". The term, coined by Steuer, includes both physically real but remote environments and virtual or simulated environments. Their common denominator is that "information is not transmitted from sender to receiver; rather mediated environments are created and then experienced" (1995, p. 37). The term "presence" has come to describe the way a mediated environment is experienced as though it were real. Biocca (1997) calls this the "illusion of being there". Yet if, as Jones (2007, p. 64) puts it, presence is the "response to a mental model of an environment that takes shape in the mind of the individual based upon a combination of cues that originate both externally and internally", how is it affected by a major sensory deficit?

People with a visual impairment can increasingly gain access to low-immersive, mediated environments such as cinema and television (TV) by means of audio description (AD). AD is a descriptive commentary that is woven between existing dialogue and critical sound effects. It provides a verbal picture for those unable to perceive visual images for themselves (Whitehead, 2005). In the UK, TV broadcasters are legally obliged to provide AD for 10% of their programmes

(Communications Act, 2003) and some voluntarily describe up to 50% (Ofcom, 2012). In a report for the Royal National Institute of Blind People (RNIB), Evans and Pearson suggest that engaging with screen-based media is considered "as vital a part of the cultural and social interactions of visually impaired audiences as it is for sighted viewers" (2009, p. 374). Yet, although extensive research into presence has been conducted with sighted audiences, very few studies have explored presence in people who are blind or partially sighted.

Biocca argues that, at any one time, individuals can feel present in one of three environments: the physical, non-mediated environment; the virtual environment; or the environment of the imagination that is dependent on the user's internally-generated imagery. This thesis presents a series of studies testing the impact of visual impairment on all three. As Lessiter et al. (2001, p. 283) point out "real-world experience is useful to presence research insofar as it serves as a benchmark, or standard, against which to subjectively judge levels of presence in mediated environments". By researching aspects of perception, experience and presence with those whose sight is impaired, light may also be shed on the relationship between presence and visual perception. The following studies represent an initial step towards understanding the psychology of AD, with implications for AD practice.

1.2 Perception: A theoretical framework

1.2.1 Models of perception

Perception can be defined as "the means by which information acquired from the environment by sense organs is transformed into experience of objects, events, sounds, tastes etc." (Roth, 1986, p. 81). It is how we engage with the world around us. Yet the nature of perception has been the subject of debate since at least the days of the Greek philosophers. Plato (427 – 347 B.C.) believed that the body and the mind were two distinct entities. He considered sensory information unreliable and insufficient by itself to lead to understanding. He argues "some accounts of things which the senses give us do not make us think, for the senses seem good enough judges of them, but others do, because sense experience gives us nothing we may put any faith in" (Richards, 1972). Plato's dualist ideas have persisted. For cognitive psychologists, perception combines bottom-up processing that is data-driven, dependent on the information available to our eyes, ears, nose, and body-based senses (kinaesthetic, efferent and inertial), with top-down processing influenced by stored knowledge (Roth & Bruce, 1995).

1.2.2 The theory of perceptual hypothesis

This interactive, reciprocal approach dates back to one of the founding fathers of psychology, Hermann von Helmholtz (1821 – 1894). He defined perception as an active way of drawing "unconscious inferences" from sensory data. In the 20th century this was extended into the theory of perceptual hypothesis (Bruner & Postman 1949; Neisser, 1976). In this view, our understanding of the world is a cognitive interaction whereby sensory data is interpreted according to hypotheses about the environment that allow an individual to infer and anticipate. For example, we hypothesise that a bottle stored in a fridge is likely to contain liquid that is suitable for drinking. Experience enables us to make finer perceptual distinctions, emphasising some aspects of the stimulus and minimising others. The shape of the bottle, the material it is made from, the colour and viscosity of its contents can lead us to make a more accurate hypothesis (this is milk) that allows us to interact more effectively with our environment (we can pour it on our cereal confidently). Experience of pouring sour milk on our cereal, however, may lead us to take further measures to identify the state of the liquid (by reading the sell-by date or sniffing the contents or both) before proceeding with the action. This cognitive approach argues that perception is the process of constructing mental representations from external stimuli. Ever-changing and unstable, stimuli are constantly compared with permanent representations or "knowledge of the world" (Roth & Bruce, 1995, p.13) stored in the brain as memory (does the smell of this particular milk resemble that of good milk or off-milk?).

1.2.3 Gibson's ecological approach

Most early studies of perception centred on the visual modality. In the 1970s, the psychologist J.J. Gibson developed what he termed "an ecological approach to visual perception" (1979). He took issue with the prevailing model of vision as a series of objects projected on the retina, like pictures projected onto a screen. Instead, Gibson argues that "natural vision depends on the eyes in the head on a body supported by the ground" (1979, p. 1). We are not passive recipients of a static, snapshot view of the world. We look around and move around while we are looking. Gibson suggests that perception is not merely a response to a stimulus but rather a way of picking up information: "Perception may or may not occur in the presence of information" (ibid. p.56).

Visual perception, therefore, is not limited to the sensory information we take in through our eyes at any given moment, but develops cumulatively. We compare what we see with what we have seen before, with what we know. As Kuhn put it, "what a man sees depends both upon what he looks at and also upon what his previous visual–conceptual experience has taught him to see" (1970, p. 113). Gibson describes visual perception as an interactive process. The brain determines what our eye is drawn to, and how we move our head or body to take in more or less of what is presented to us. The visual information we receive is, in turn, processed by the brain, resulting in physiological changes (e.g. the pupil dilating or retracting) or task-driven changes (e.g. we like or dislike what we see) and results in the eye, head or body moving in to look more closely or turning away.

Crucially, Gibson argues that objects and environments are not neutral. They are not simply there for us to respond to, as we will. Objects afford action. An object of a particular size, say a twig, affords us the opportunity to pick it up, perhaps to use as a tool, or as fuel for a fire. A larger object, the tree branch for example, may afford grasping but not carrying. An attached object, the tree itself, cannot be carried and nor, if the trunk is wide, can it be grasped. However, it may afford climbing. If it is a fruit tree, the fruit may afford eating. Recognising those affordances is the purpose of perception. Sight in particular enables us to move through an environment safely, engaging with that environment and the objects it contains. "Perceiving," in Gibson's view, "is an act, not a response…an achievement, not a reflex" (ibid. p. 149).

1.2.4 Attention

The sheer volume of incoming sensory data, from our eyes, ears, nose, tongue, skin, muscle movement and linear and angular acceleration of the body, needs to be filtered in some way if we are not to suffer from cognitive overload. Attention “selects” what is of relevance, allowing the brain to neglect the rest (Koch & Tsuchiya, 2006). Selection may be based on exogenous, bottom-up factors such as visual flicker, colour, luminance or movement and top-down, endogenous factors that are related to the needs or interests of the individual. Such task-driven factors can override saliency or boost it artificially (Henderson et al, 2009; Torralba et al., 2006). Smith (2011) used eye-tracking to demonstrate that viewers shown scenes of famous Edinburgh landmarks were drawn to salient elements (e.g. a girl in a bright pink coat). Asked to identify the landmarks, however, they neglected the little girl and directed their gaze to the landmark in question.

Attention may or may not be conscious, i.e. we may or may not be able to remember and report on particular features of a scene. For example, affordances draw our attention, even if we are not consciously aware of them. When shown a target with a graspable handle on either the left or right hand side, observers respond more quickly and more accurately if the response hand is on the same side as the handle, than if it is on the opposite side (Tucker & Ellis, 1998 cited in Murphy et al., 2012). This holds true even if participants observe a photograph, rather than an object that could be physically grasped, although in Murphy, van Velsen and de Fockert’s study the advantage was only maintained under low cognitive load. Ahmed & de Fockert (2012) have found that attention is more easily distracted in conditions of high cognitive load and in individuals who have low working memory capacity. This is in line with the load theory of attention (Lavie et al., 2004) which proposes that strain on working memory reduces ability to focus on the task in hand, making it harder to distinguish between relevant and irrelevant stimuli. Load is explored in more detail in Chapter 2.

Attentional cues are not restricted to the visual modality but occur crossmodally. Exploring spatial attention, Driver & Spence (2004) found that a task-irrelevant but salient stimulus in one modality (touch, audition or vision) could attract covert attention in another. However, Santangelo and Spence (2008) suggest that, in the laboratory at least, bimodal (auditory/visual) cues are more resistant to manipulation of perceptual load (such as increasing the number of distractor stimuli). They argue that more perceptually demanding tasks “raise the threshold” which peripheral stimuli must overcome to deflect spatial attention. As with unimodal processing, they claim that for multimodal cues, exogenous attention is not “automatic” as previously assumed but can be overridden by task-driven and volitional factors. In all, it seems that perception depends on a continuously shifting balance between competing demands of external stimuli and internal resources. Knowledge and experience help screen unnecessary or redundant data, freeing up as many cognitive resources as possible in pursuit of particular goals.

1.2.5 Cultural factors

Cognitive anthropologists recognise that perception is culturally determined (Thagard, 2012). The ability to discriminate colour, for example, is affected by linguistic categorisation that varies from one culture to another (e.g. Winawer et al., 2007; Goldstein, Davidoff & Roberson, 2009). Cultural differences are not necessarily restricted to the visual modality. Testing visual-flavour matches with the Himba of Northern Namibia, for example, Bremner et al. (2012) found that unlike Western participants, the Himba did not preferentially map spiky/rounded shapes to samples of carbonated/still water, and mapped bitter chocolate to a rounded shape, in the opposite direction to that preferentially chosen by Westerners (Ngo et al., 2011). While the chocolate-shape choices could be explained by the sounds of the Himba words used to denote bitter and sweet, Bremner and his colleagues put the water-shape mapping down to visual conditioning. In the West, for example, brands of carbonated drinks often feature angular motifs (e.g. Spence, 2012; Spence & Gallace, 2011).

This idea of cultural influence on mental processes can be extended such that subgroups within the same culture may develop different linguistic, bimodal and multimodal associations. In particular, the lack of a major sensory modality i.e. vision, may affect not only *what* information is perceived but also *how* that information is processed.

1.3 Visual processing and visual impairment

1.3.1 Defining visual impairment

Literature relating to studies with blind and partially sighted (B&PS) people uses an array of terminology to discuss those with a visual impairment (for a review see Cattaneo & Vecchi, 2011). In the UK, the legal definition of blindness is “a person is so blind that they cannot do any work for which eyesight is essential” (http://www.nhs.uk). This is regarded as between zero and 10% of nominal[1] visual acuity. This legal definition fails to distinguish between those who see nothing at all (totally blind) and those with severe visual impairment who have light perception and can, at least, perceive the difference between light and shade. Blindness encompasses not only *extremely poor* visual acuity (less than 3/60 i.e. the inability to see at 3 metres what a person with "nominal" visual acuity can see at 60 metres); but also *poor* visual acuity (3/60 – 6/60) combined with *severe* reduction in the visual field; and *average* visual acuity (6/60 or better) with *extreme* reduction in the visual field. Visual impairment also refers to people with partial sight which is legally defined as *very poor* visual acuity (3/60 - 6/60) combined with a full field of vision; *moderate* visual acuity (up to 6/24) and a *reduced* field of vision or blurred or cloudy central vision; or relatively good visual acuity (up to 6/18) but *severe* reduction in the visual field.

Another aspect of this broad spectrum is that only a very small number of people are born with no sight or lose it in early infancy, such that they have no visual experience. More commonly, sight problems develop later in life, through accident or illness. Some eye conditions, such as retinitis pigmentosa, deteriorate so the impairment progresses from partial sight to total blindness. A few conditions, such as congenital cataracts, are reversible so initial blindness is replaced by sight. Some acquired conditions can be reversed to a certain extent e.g. diabetic retinopathy can be improved by a kidney transplant. Some of the

---

[1] "nominal" visual acuity = 6/6 (formerly 20/20); acuity of 3/6 is considered half as good; acuity of 12/6 is twice as good.

variety of visual experience within the B&PS population is illustrated schematically in Fig. 1.1.

Most B&PS people are elderly. Of those registered blind in England (NHS, 2011), 66% are aged 75 or over, and only 0.48% are aged 4 or under. The figures are similar for those who are registered with partial sight: 64% aged 75 or over; 0.35% aged 4 or under. Overall, as vision problems are mostly age-related, and given the ageing population, the number of B&PS people in the UK is expected to increase.

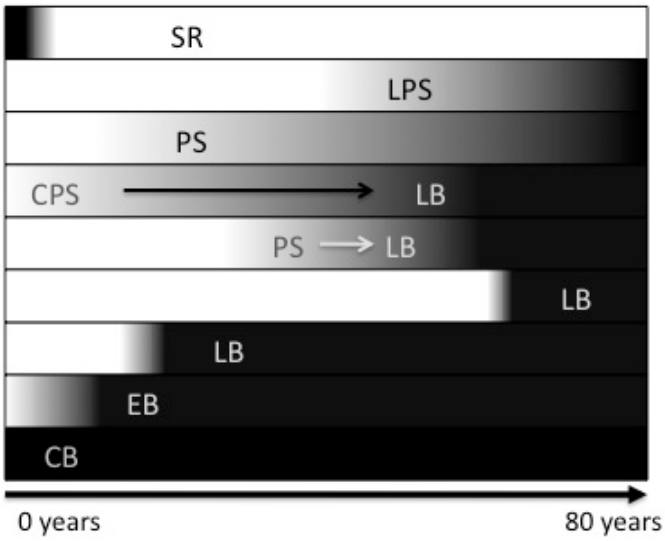


Fig. 1.1 A schematic representation of the diversity of visual experience within the blind and partially sighted population. CB = congenitally blind; EB = early blind; LB = late blind; PS = partially sighted; LPS = late partially sighted CPS = congenitally partially sighted; SR = sight restored. Darker shading equates to more profound visual impairment.

1.3.2 Models of visual processing

1.3.2.1 The central visual pathway

In order to discuss the impact of visual impairment, it is important to have some understanding of sight. At its most basic, the visual processing system is triggered by light entering the brain via photoreceptors at the back of the eyeball. Visual information is conveyed along the optic nerve leading from each eye to the optic chiasm where the optic nerves converge. One axon from each crosses over into the contralateral optic tract. From this point the optic tracts diverge, pass round the mid-brain (cerebral peduncle) and feed in to the lateral geniculate nucleus (LGN) of the thalamus. The LGN axons radiate from here, through white matter of the brain, to the primary visual cortex (striate cortex or V1) of the occipital lobe at the back of the brain. The rest of the occipital lobe makes up the visual association (extrastriate) cortex, concerned with recognition, colour, motion and depth (Crossman & Neary, 2010).

1.3.2.2 Dorsal and ventral streams

Ungerleider, Mishkin and Macko (1983) proposed that information coming from the extrastriate cortex is separated into two processing streams: the dorsal stream (the “where” pathway) carrying details of motion, colour and spatial features to the parietal lobe; and the ventral stream (the “what” pathway) concerned with object recognition, faces, shape, size and text that connects to the medial temporal lobe. Goodale and Milner (1992) refined the model, suggesting that each pathway was used for a different task: the dorsal pathway for “action” (as it feeds into the motor cortex of the frontal lobe) and the ventral pathway for “perception”, combining visual recognition with memory and emotion (as it feeds into the hippocampus and amygdala). They suggested that, while the ventral stream elicits a conscious percept, the dorsal stream does not directly connect to consciousness.

1.3.3 Visual "maps" and crossmodal integration

This unimodal approach to cognition, suggesting discrete processing pathways for different sensory inputs, is being revised in the light of increasing evidence of crossmodal integration. Laurienti et al. (2002), using functional Magnetic Resonance Imaging (fMRI), showed that activity in the visual cortex was "switched off" by novel auditory information and that novel visual information similarly inhibited activity in the auditory cortex. Beer et al. (2011) have subsequently provided evidence from diffusion tensor imaging to show the existence of direct white matter connections between visual and auditory cortex. They argue that these links suggest multisensory integration occurs much earlier in the processing system than previously thought. de Haan and Cowey (2011) have put forward a new model replacing visual pathways (with their implication of a hierarchical progression) with a "flatter" system based on visual maps. They define a visual map as "a retinotopic area in the posterior brain that is concerned with the processing of (a) specific visual cue(s)" (2011, p. 460). Such areas can be recruited, patchwork fashion, according to the task in hand. Numerous maps may be used in conjunction to perceive and identify a single characteristic. Presumably such maps could also recruit areas that process sensory information from multiple modalities.

1.3.4 Cognitive processing in the absence of sight

Vision is only one sensory source that must be integrated with non-visual modalities to produce a coherent representation of the world. Spence and Gallace (2011b) argue that human perception is inherently multisensory but that sight predominates. Studies have shown that vision overrides information pick-up from other non-visual modalities such as touch (Hartcher-O'Brien et al., 2010) and hearing. When presented with bimodal (auditory and visual) stimuli, for example, sighted people respond more quickly to the visual component than the auditory one (Koppen & Spence, 2007). This finding is not new. McGurk showed that when watching a woman's lips repeatedly pronouncing the syllable 'ga', while listening to the syllable 'ba', participants reported hearing the syllable 'da'. This "McGurk effect" holds true, even when we are conscious of it (McGurk & MacDonald, 1976; Alsius et al., 2005). The art of ventriloquism also relies on prioritising the visual stimulus of the dummy's moving mouth over the auditory stimulus (the words coming from the ventriloquist), such that the dummy appears to be speaking. We experience a similar effect every time we go to the cinema, attributing words to the characters on screen in front of us, even though the voices are relayed via speakers placed around the side and back walls of the auditorium (Murray et al., 2005).

Vision is generally acknowledged to provide the main reference frame for organising incoming sensory data (Rock, 1985; Putzar, Goerendt et al., 2007). The question then arises: what happens to cognitive processing in the absence of sight? In particular, what are the developmental outcomes for congenitally blind people (CB) i.e. people blind from birth? Röder and Rösler (2004) outline 3 possibilities:

i. If visual experience is necessary for normal cognitive processing, CB will show impaired performance in all modalities.
ii. If senses develop independently, CB will show impaired performance

only in the visual modality.

iii. If non-visual senses compensate for lack of vision, CB will out-perform sighted people in tasks involving non-visual modalities.

Cattaneo & Vecchi (2011) argue that differences arising from varying types of visual impairment can inform the debate on models of sensory processing. For example, Heller, Wilson et al. (2003) showed that late blind people (LB) were better than early blind (EB) or blindfolded sighted participants at identifying a target figure from a line drawing. They put this down to a combination of visual experience combining with haptic practice. They argue that such results extend Röder and Rösler's list of possibilities as follows:

iv. If visual experience is necessary for normal cognitive processing, then CB will show impaired performance in all modalities compared to their EB/LB and partially sighted (PS) peers.

v. If senses develop independently, CB/EB will show impaired performance compared to their LB/PS peers only in tasks requiring visual memory.

What constitutes "Late" as opposed to "Early" blindness varies across studies. Cattaneo & Vecchi (2011) review 44 studies in which the cut-off ranges from 2 – 7 years[2]. This difference is important as a 7 year old is more likely than a 2 year old to have retained some degree of visual memory. The brain has also been shown to be highly plastic, especially in the young (e.g. Theoret,

[2] The cut-off chosen for the studies in this thesis is 3 years. See section 2.4

Merabet & Pascual-Leone, 2004; Huttenlocher, 2009) allowing different cortical connections to develop (Shu et al., 2009). Again, the age of cut-off is debated but there is some agreement that the phase of greatest brain plasticity ends by the age of 14-16 years (Wan et al., 2010). This leads to the following possibilities:

vi. If only CB and EB benefit from cortical reorganisation, they will outperform LB/PS in tasks involving non-visual modalities

vii. If compensatory plasticity continues throughout life, difference in performance between CB, EB, LB & PS will vary in accordance with extent and duration of sight loss.

viii. Alternatively, experience may compensate for lack of early brain plasticity, such that performance between CB, EB, LB & PS appears similar but is the result of different strategies or processing systems.

1.3.5 Sensory compensation

Evidence for the sensory compensation hypothesis (see iii above) is inconsistent. In the tactile modality, for example, EB *and* LB were better at a grating-detection task than sighted people (Goldreich & Kanics, 2003; 2006). Yet, EB have been shown to have a compensatory advantage in some aspects of auditory perception, including pitch discrimination (Gougoux et al., 2004) and more efficient processing of simple auditory stimuli (Stevens & Weaver, 2009). Investigating the ability to localise sounds, Voss, Lassonde et al. (2004) found no difference in performance of EB/LB and sighted people (S) in central space, but enhanced performance for EB/LB for sounds in peripheral space. However, in another experiment comparing localisation of sounds in the vertical plane, B were outperformed by S (Lewald, 2002).

This suggests sensory compensation deriving from functional reorganisation of the sensory cortices is driven by both intra-modal brain plasticity and by experience that encourages new connections to develop between modalities. For example fMRI data (Fiehler & Rösler, 2010) for CB and blindfolded S executing active and passive hand movements showed overlapping patterns in areas of the dorsal stream, but CB participants also showed some activation in the extrastriate cortex and the auditory cortex. The extent of this was linked to the age at which they had received orientation and mobility (O&M) training. Occipital recruitment by non-visual modalities is not restricted to CB/EB and has been reported in studies with LB participants. For example, Burton, Snyder et al. (2002) reported activity in the visual cortex during a Braille reading task by both CB and LB participants. They put forward the explanation that, although in sighted people the visual cortex is suppressed during a tactile task (possibly to avoid cognitive overload), this is evidently not necessary in a person with no sight.

Collignon, Voss et al. (2009) point out that inconsistent evidence may arise for a number of reasons in addition to those described above. As mentioned, the sight characteristics (aetiology, age at onset etc.) of blind participants may vary from one study to the next. This is exacerbated by the generally small sample size in studies involving blind participants. In addition participants may be using processing strategies that vary both within and between groups. Perhaps most importantly, the task that forms the basis of the experiment may be more relevant to one group of participants than another. This is returned to below.

1.3.6 Attention and visual impairment

Röder, Kramer & Lange (2007) suggest blind people benefit from enhanced attention rather than enhanced non-visual senses per se. Required to attend to an auditory offset stimulus presented at a short or long interval, and in the left or right hemi-field, they showed that CB participants attended to both spatial and temporal inputs, while sighted people only attended to spatial inputs. Other studies have shown that blind people are better at divided attention tasks in response to auditory and tactile stimuli. Occelli, Spence, and Zampini (2008) for example found that EB and LB were better than S at discriminating the spatial location of simultaneously presented auditory and tactile cues. CB participants were also better at a selective attention task in which participants had to judge the number of tactile stimuli (single touches) when presented together with irrelevant (multiple) auditory tones (Hötting & Röder, 2004).

1.3.7 Crossmodal integration and visual impairment

These attentional advantages have been cited as evidence that blind people are less good at multisensory integration (Hötting et al., 2004). Shu et al. (2009) used diffusion tensor tractography to show a decreased degree of connectivity in the cortical networks of CB, especially in the visual cortex. However, other areas relating to motor and somatosensory function showed the opposite. Behavioural evidence also suggests that isolation of modalities in blind people is not clear-cut.

Eardley and Pring (2011) found that blind and sighted people were equally affected by a multisensory distraction task: ratings of vividness of taste for a variety of familiar foods decreased while sucking a sweet.

Röder, Focker et al. (2008) put forward the idea that crossmodal integration is only possible in the presence of an external reference frame, namely sight. They used a crossed-hands paradigm to show that S made fewer successful temporal order judgments in response to tactile stimuli in the crossed-hands position, while CB were unaffected. However, visual dominance in visuo-tactile tasks has been demonstrated in, for example, the rubber hand illusion (Botvinick & Cohen, 1998). An individual who watches a dummy hand being stroked, while their own unseen hand is stroked in synchrony, comes to believe that the rubber hand is their own. This has been demonstrated not only with a dummy hand physically present but also with a virtual hand (IJsselsteijn et al., 2006). Given the strength of visual dominance, then, it seems hardly surprising that S are affected by the visual element of a visuo-tactile cue, but CB are not. While Röder et al.'s study makes the case for multisensory integration in sighted people, it does not prove the lack of multisensory integration in blind people. Collignon, Voss et al. (2009) used a more task-appropriate paradigm, asking participants to respond to a bimodal (auditory or tactile) stimulus presented on the left or right side, by pressing a button with their left or right hand. The crossed-hands position caused difficulties for S and LB in response to both modalities, but only in response to the sound stimulus for CB. They argued that CB rely on sound to create an organisational, external reference frame in the way that S (and LB) rely on vision (or visual memory).

Further mixed evidence for multisensory integration resulting from brain plasticity comes from research with people born with binocular cataracts, who have them removed in childhood. After an initial period of visual deprivation in infancy, then, sight is able to develop. Putzar et al. (2007) reasoned that if such newly acquired sight developed normally, these individuals would show the benefits of bimodal (audio-visual) stimuli. This proved not to be the case for a speech perception task. Sighted participants gained from adding the visual stimulus to the auditory one; previously-blind participants did not, despite having had sight for at least 14 years at the time of the experiment. However, in a more recent study (Putzar et al., 2012) reaction times to auditory–tactile, auditory–visual, and tactile–visual stimuli were similar between sight groups. The researchers argue that some multisensory responses (and not others) can be developed by experience.

1.3.8 Mental imagery and visual impairment

Over the past century or more, research into the nature of sensory processing has explored the role of mental imagery (e.g. Galton, 1880; Betts, 1909). As Thomas (1999, p. 208) expresses it, imagery is “a quasi-perceptual experience; experience that significantly resembles perceptual experience (in any sense mode), but which occurs in the absence of appropriate external stimuli for the relevant perception”. Kerr (1983) points out that researchers have a tendency to use visual terms such as “look” and “see” when exploring mental imagery. By drawing visual analogies such as “scanning” with the “mind’s eye”, she suggests they effectively rule out blind people’s experience. Campos (2004) suggests that by drawing on non-visual sensory modalities, blind people construct mental images that they rate as vividly as mental image ratings of sighted people, although such images will lack specifically visual properties such as colour or brightness.

Regardless of the nature of such images, there is certainly a functional similarity. In a comprehensive literature review, Cattaneo et al. (2008, p. 1347) found "ample evidence to indicate that the performance of blind individuals is remarkably similar to that of sighted individuals in tasks presumed to involve visual imagery". Cattaneo and Vecchi's own experiments (e.g. 2006) make a distinction between people with acquired sight loss and those blind from birth. They suggest the latter, including those born with monocular vision, perform differently from both S and LB on spatial imagery tasks involving perspective and mental manipulation of objects in 3 dimensions. Vanlierde & Wanet-Defalque (2005) also showed that LB performed similarly to S in mental imagery experiments estimating distance of objects, compared with participants who had no visual memory. By contrast, CB were not able to imagine an object so close that it "overflowed" the boundaries of their mental image. This suggests that visual experience has a direct impact on imagery.

However, just as vividness of imagery varies amongst sighted people (e.g. Richardson, 1999) the same may be true for those with a visual impairment. This possibility has rarely been considered in the literature, but is described by Oliver Sacks in an article for the New Yorker (2003). Inspired by the autobiography of John Hull (1997), who gradually lost his sight through retinitis pigmentosa, Sacks assumed all late blind people resigned themselves to living without sight by gradually giving up their interest in the visual world. However, contact with other blind people showed that some actively developed their desire and ability to maintain and manipulate visual imagery, even if it led to wildly inaccurate visualisations. This led him to doubt the existence of “a ‘typical’ blind experience” (p. 54). Sacks reported the comments of two LB people who spoke not only of a heightened visual imagination but also “a much readier transference of information from verbal description – or from their own sense of touch, movement, hearing or smell – into a visual form”. Sacks concludes that, while for one person the ability to visualise is critical to function effectively, for another its absence is not inhibitory. Ultimately, he argues, it is difficult to decide what in the mental landscape is purely visual or auditory or linguistic, or emotional: “they are all fused together and imbued with our own individual perspectives and values” (p. 59).

1.3.9 Amodal, multimodal and supramodal imagery

To explore the relationship between perception and imagery, Struiksma et al. (2009) have investigated the links between language and spatial imagery that they claim “is different from sensory imagery in that it is not bound to a single input modality” (p. 146). Leaving this claim aside for the moment, they argue that language provides an alternative to sensory perception as it consists of abstract information. They give the example of going to buy a pineapple from a shop. This can be achieved either by conjuring up a mental image of a pineapple which they define as “a visual image containing the shape and form of a pineapple; a sense of the tactile information of the texture one feels when grasping the fruit; olfactory information about its smell; a quasi motor program containing movement and directional information on how to walk towards the fruit stall" (2009, p. 145) or by acting on a set of verbal instructions, processed in working memory. In this way spatial imagery benefits from dual encoding: spatial language can trigger a "pictorial" image or purely propositional information.

Language has been variously thought of as amodal (e.g. Fauconnier, 1999) i.e. not directly linked to any specific modality, or multimodal (e.g. Barsalou et al., 2008) such that it is grounded in experience and leads to simulated perceptual experience, processed by the modality-specific cortices. As cited above, such simulated experience is a common definition of mental imagery. Struiksma and her colleagues opt for a supramodal model of imagery whereby the image is fed by both modal information and amodal linguistic information, thus activating modality-specific cortices and a cortical region devoted purely to spatial information. All sensory modalities may contribute, but the weighting will vary according to importance, being predominantly visual in sighted people. This weighting will be different for those with a visual impairment. The researchers reason that, given the sensory compensation hypothesis, the missing visual stream is counterbalanced by increased information from motor, auditory and tactile modalities.

1.4 Presenting a "linguistic compensation" hypothesis

Two questions arise from Struiksma et al.'s proposition. The first is whether they are correct in their assumption that spatial imagery is unique in combining information from multiple modalities. Given increasing evidence for crossmodal integration (discussed above) it seems unlikely that imagery in any modality could be experienced in isolation. The hypothesis that all mental imagery is multimodal or supramodal will be explored in Chapter 5.

Secondly, if the weighting of modality-specific information can change to provide compensation, so, presumably, could the weighting of linguistic information. For example, in Putzar et al.'s (2007) experiment mentioned above, people who regained their sight following operations for congenital cataracts did not benefit from a bimodal advantage when visual cues were added to speech. The researchers put this down to such individuals missing a critical developmental window for brain plasticity and thus failing to develop a multisensory response. However there is consistent evidence from neuroimaging studies that speech is processed in the occipital cortex of blind but not sighted people (Bedny et al., 2011). Amedi et al. (2004), for example, showed that repetitive transcranial magnetic stimulation (rTMS) of the occipital pole of B (but not S) participants reduced the accuracy of verb generation. Ricciardi & Pietrini (2011) point to a number of studies that have found evidence for crossmodal neuroplasticity in what they call the "blind brain" by which the occipital cortex is recruited not only for memory and word-based tasks but also for perceptual tasks including auditory-spatial processing (Collignon et al., 2011) and auditory discrimination (see Merabet & Pascual-Leone, 2010, for a review). Conversely, lesions of that brain region lead to difficulties not only in generating words but also give rise to perceptual problems such as tactile sensitivity (Noppeney, 2007). This suggests that, in blind people, perceptual and linguistic processing maybe intrinsically linked. If the brain is plastic enough to rewire crossmodal connections, is it too far-fetched to suggest that it might do the same with language-processing, such that language replaces sight as an external reference frame?

There is an experiential argument as to why this might be the case. An ongoing study into the development of blind infants (Dale, 2013) shows that, at around the age of 18 months when a normally-developing child is beginning to explore the world by crawling and reaching out for objects, a blind baby will sit motionless, unless another individual interacts. This interaction is likely to take the form of verbal encouragement on the part of the parent, who will simultaneously present an object for the infant to explore by sound and touch, or encourage the infant to move towards them. For sighted infants, congruent information from different modalities can be matched implicitly: a sighted baby hearing a mewing sound can turn its head to see a cat. In the next encounter, on seeing a cat, the child will expect to hear it mew, or put out their hand to stroke it. A blind child relies on being told a cat is present to build up the same auditory or tactile expectation. Similarly, for a sighted child, much of the shared experience of the external world can remain implicit, relying on pointing or eye-gaze. Shared interaction for the blind child necessarily involves language.

Even for an adult with sight loss, everyday interactions are likely to be accompanied by a greater exchange of verbal information than is usual between those who can perceive that information for themselves. When bringing a sighted person a cup of tea, for example, the tray may be simply set down with an instruction to “help yourself”. For a blind person, the arrival of the tray will be accompanied by a description of where the tray is, how many cups are on it; how full a particular cup is; whether or not milk has been added; which way the handle is facing etc. Arguably, then, in the absence of sight, speech is essential to the integration of sensory input and therefore to perception. In this sense, speech replaces visual input. This “linguistic compensation” hypothesis will be tested in the studies described in the following chapters.

1.5 Summary

After a general introduction to perception, this Chapter has discussed what is currently known about the cortical pathways involved in visual perception, multisensory integration and imagery. It has introduced current theories of how the sensory processing system is affected by visual impairment. It has outlined the sensory compensation debate, and explored what behavioural and neuroimaging studies tell us about the brain in people with varying sight characteristics. It has also introduced an alternative, or complementary hypothesis of linguistic compensation. How visual impairment impacts on engagement with mediated environments and, in particular, sense of presence is the subject of the next Chapter.

Chapter 2 Introduction: Presence, Visual Impairment and Audio Description

2.1 Introduction

Can B&PS people experience presence in a mediated environment, given that their access to the visual information stream is reduced or absent? This Chapter outlines theories of presence and models the way in which sight characteristics might impact on access to AV media. It describes the methodologies used in this research and defines the questions in relation to visual impairment, perception, experience and presence that form the basis for the studies presented in subsequent chapters.

2.2 What is presence?

The compound word "telepresence" was put together over thirty years ago to describe human interaction with remote-access technology (Minksy, 1980). Since then the concept of presence has been extended to embrace the psychological sense of immersion and engagement in any mediated environment, such that it appears unmediated (Lombard & Ditton, 1997).

Waterworth & Waterworth (2003a) raise the objection that immersion and engagement are two separate constructs. In their view, presence is a biological phenomenon experienced by a conscious being (i.e. humans) immersed in the external world. They argue that presence is the means by which reality can be distinguished from illusion. It is separate from an internally-constructed, imaginal state because engagement in the imaginal world necessitates disengagement from physical cues in the "here and now". In contrast to presence, Waterworth & Waterworth term this deliberate disengagement "absence". They argue that we cannot feel conscious of our own bodies in the external or virtual world *and* lost in our own thoughts at the same time. In their Focus, Locus and Sensus model (2001), they draw on the image of a light bulb swinging between two rooms. It can illuminate the internal world or the external world but not both simultaneously.

Biocca (2003) proposes a more fluid model, with the user's experience of presence constantly shifting between the virtual space, the physical space and the space within their own imagination. Yet Jones (2007) points out that both models are based on an unwritten assumption that conscious attention must be directed to one or other of these spaces. Attention is itself part of the cognitive process, that, as we have seen, is affected by the nature of the stimuli and the experience and motivations of the interpreter. To extend the image of the light bulb, an individual's interests, knowledge and perceptual systems would determine the bulb's degree of luminosity and the speed and angle of its swing. In a Gibsonian view, any separation between external and internal world breaks down completely such that, presence is simply the sense of being in an environment whichever type of environment that might be.

2.2.1 Presence and consciousness

Conscious experience may be triggered by stimuli in the phenomenal world but can also be triggered in other ways e.g. pathologically, in hallucinations, or artificially by visual and auditory cues arising from mediated environments such as 3D cinema, holograms and virtual environments (VEs). If we accept this is the case, Velmans (1999; 2008) suggests that the problem of how mediated environments are experienced as real then drops away. When sensory inputs "provide information which resembles that arriving from actual objects in the world [the] mind/brain models this information in the normal way and constructs what it normally constructs when it receives such input – a perceived, phenomenal world located and extended in the three-dimensional space beyond the body surface" (p.22).

How natural the VE appears to be (also known as Ecological validity) is one dimension of presence that has been identified together with Spatial presence, Engagement or involvement in the mediated world, and Negative effects such as headaches or eyestrain that may result from interaction with the VE (Lessiter et al., 2001). The VE may also include people. Biocca uses the term *Social* presence to describe the illusion of being with a mediated person. Levels of presence have been shown to increase in line with "the degree to which a user feels access to the intelligence, intentions, and sensory impressions of another" (1997, p. 22). IJsselsteijn, Freeman and de Ridder (2001) have also identified *co-*presence that is a combination of Spatial and Social presence.

2.2.2 Determinants of presence

Just as some people experience hallucinations and others do not, the question that has fascinated presence researchers is *why* some people more than others neglect perceptual information about the here and now in favour of the mediated environment? What influences our engagement with this *constructed* reality (Shapiro & McDonald, 1992) such that we experience it as real? What are the determinants of presence?

2.2.2.1 Media Form

Lombard (2008) has recently argued for a return to use of the term "telepresence", originally coined to distinguish "mediated" presence from "natural" presence (Steuer, 1992), in order to emphasise the role of technology. VEs were initially defined in relation to the sophistication of their hardware i.e. media form. Sheridan (1992) suggests this can influence presence in three ways: the range of sensory stimuli presented to the user; the extent to which a user can control their own receptors (e.g. adjusting their view of the environment) and how much a user can modify what is presented to them.

In line with Velmans' model, early researchers mistakenly assumed a direct, linear relationship between presence and more immersive media forms. Zelter (1992), for example, proposed that the greater the number and fidelity of sensory input and output channels from a display, the more immersive the technology and the greater the experience of presence. This cybernetic definition (Draper, Kaber & Usher, 1998) has since been shown to be simplistic. Even low-immersion media such as TV or books can induce a sense of presence (e.g. Bracken et al., 2010; Bracken, 2005; Jones, 2007; Wirth et al., 2007). When reading a book, for instance, the ability of text to evoke presence suggests that higher-order mental processes involving cognition and imagination may compensate for any lack of direct sensory experience resulting from media form (Gysbers et al., 2004). However, Gysbers and his colleagues point out that simply being able to visualise or imagine the environment is not sufficient to create a sense of presence; the reader must regard themselves as “being there” (Barfield et al., 1995). This depends on the reader’s motivation and, perhaps, their willingness to focus on textual cues that confirm they are in the imaginal environment, while suppressing perceptual cues from their real environment. In short, they must select the mediated environment as their primary egocentric reference frame (PERF, see Ricke & von der Heyde, 2002).

2.2.2.2 Individual differences

In their review of presence literature, Sacau, Laarni and Hartmann (2008) agree that levels of presence are governed as much by qualities of the individual as of the media that trigger it. Among the many contributory elements are personality, cognitive style and ability, age, gender, imagination (Heeter, 1992; Jurnet, Beciu & Maldonado, 2005), the ability to suspend disbelief (Lombard & Ditton, 1997), attention (Carroll, 1993), affect and emotion recognition (Västfjäll, 2003) and deficits in cognition and perception (for a review see Stanney, Mourant & Kennedy, 1998). Steuer (1993) points out that, even for deeply immersive VEs, any definition of virtual reality must include user experience. Individuals both influence and are influenced by the hardware such that presence in a VE “is a function of both technology and perceiver” (1993, p.10).

2.2.2.3 Media content

Together with media form and user characteristics, IJsselsteijn et al. (2000)

include media content as a further determinant of presence. Content refers to the general theme or narrative depicted (Lessiter et al., 2001). Green & Brock (2002) coined the term "transportation" to describe the process by which we suspend information about the real world and plunge instead into the world of the narrative. For written texts they define this as "the experience of cognitive, affective and imagery involvement in a narrative" (2004, p. 311).

The influence of content on presence has been shown to be independent of media form. Research by Dillon (2006), for example, showed that emotive movie clips produced higher subjective ratings of presence than neutral clips, regardless of whether they were viewed on a small or large screen. Similarly, footage showing a relaxing boat ride induced a stronger feeling of presence than that of a rally-driving sequence (Dillon et al., 2001). Wirth et al. (2007), citing Kelso, Weyrauch, & Bates, (1993) and Klimmt & Vorderer (2003), argue that narrative, drama and plot all help sustain interest, and therefore presence, by encouraging the individual to select the mediated environment as their PERF. Nunez & Blake (2006) showed that Spatial presence was predicted by an individual's familiarity with, and interest in, the content.

2.2.2.4 Flow, presence and cognitive load

Jones (2007) points out that in addition to the subject matter, style factors that alter the way content is presented also affect presence i.e. "techniques and conventions germane to the medium". For example, in a study by Kraft (1986) participants found fast-cut film sequences to be more exciting compared with the same sequence filmed as one continuous take. The way in which images, sounds and text combine and progress from one to the next is called “flow”, a term borrowed from Csikszentmihalyi (e.g. 1991) who used it to express that sense of perfect concentration that allows us to be unquestioningly “in the moment” or “in the zone”.

In Csikszentmihalyi’s model, flow can only arise when there is an equal balance between the demands of the task and the skills of the person performing it: too difficult a task leads to conscious awareness through anxiety; too easy a task leads to boredom and apathy. In a mediated environment, a sequence of sounds and images that are disjointed or incomplete may lead to the former; too explicit, may lead to the latter. This can be explained through cognitive load theory as originally conceived by Miller (1956) and developed, largely in relation to learning objectives, by Sweller (1998). Ng et al. (2013) argue that a dynamic medium (a film compared with static images, for example) is essentially transient, constantly presenting the viewer with new information. This needs to be processed while, at the same time, information it has superceded must be held in working memory (WM) long enough that connections between the two information states may be made. WM must not be overstretched and, according to cognitive load theory, the perfect balance (germane load) is dependent on intrinsic load (the complexity of the message), extrinsic load (the complexity of the medium or the way the message is communicated) and the expertise of the individual. This is illustrated in Fig. 2.1.

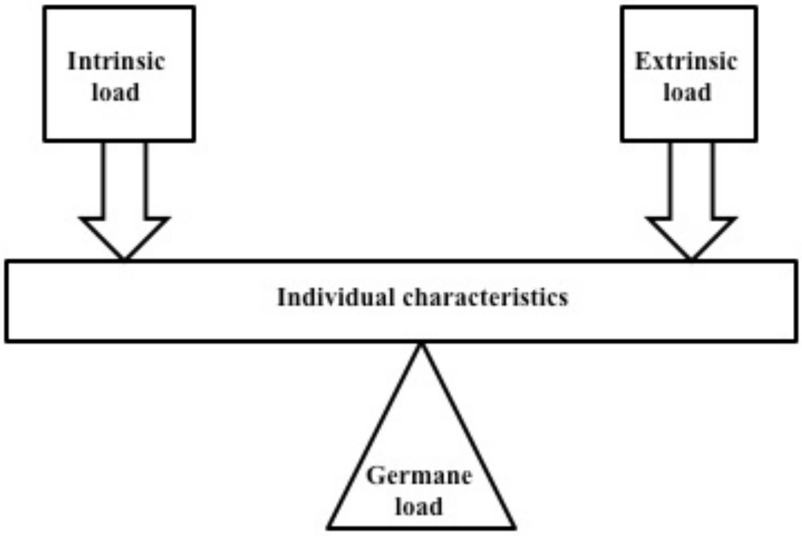


Fig. 2.1 Model of cognitive load: individual characteristics (e.g. familiarity, expertise) will dictate how much load a person's WM can bear. Too much intrinsic or extrinsic load will upset the balance needed for germane load.

Ng et al. (2013) argue that novel information can quickly overwhelm WM, whereas familiar information can be dealt with via schemas within long-term memory that is virtually limitless (e.g. Sweller, 2010). Schemas are discussed in more detail in the next section.

Schrader and Bastiaens (2012), comparing presence and learning outcomes for computer games in a high-immersive or low-immersive environment, found that high presence reduced learning (measured by comprehension and recall). They suggest that presence inhibits learning by requiring that state of augmented concentration which Csikszentmihalyi terms "flow". In this sense, flow means we are unaware of ourselves in the real environment and instead submit ourselves to the trajectory of the narrative. Arguably, in an entertainment environment, learning is not a priority.

However, it is also possible that presence may influence cognitive load positively, by increasing engagement with the *intrinsic* load leading to greater attention, and by focussing that attention on essential elements of information, thereby reducing *extrinsic* load. Glenberg (1997) suggests that an optimal combination of flow and trajectory means that fidelity of a mediated environment to every aspect of the "real" physical environment is unnecessary because, in life, we pay little attention to physical detail. For example, in walking up the path to our house, we attend only to limited sensory information augmented by stored knowledge or memory so that, from experience, we can be certain that this is our own path and not another. Both types of information are used to support action in the environment. Glenberg calls this "embodied perception" and acknowledges its similarity to Gibson's idea of affordances. Noticing what an environment affords rather than its qualities (appearance) extends to objects too. Zahorik & Jenison (1998) use the example of a basketball, arguing that in artificial intelligence applications it is recognised most easily not by the features "round", "orange" and "rubber" but rather by the fact it can be thrown, rolled or bounced. There are parallels here with Heidegger's concept of "throwness" (1962; 1977): as we are constantly thrown into situations that require action, we rarely have the opportunity for analytic (objective) reflection. We do, rather than think. Glenberg argues that, so too, in mediated environments, "we develop an action-based understanding of the situation" (1997, p. 42).

2.2.2.5 Presence, experience and schemata

Schubert et al. (2001) propose a "potential action coding theory of presence" built around a mental model, constructed both from bottom-up information (media form and content) received directly through the senses and top-down processing dependent on pre-existing knowledge (user experience). Jones (2007, p. 8) cites Jakob von Uexkull's concept of the *umwelt* as relevant here, i.e. "the mass of knowledge that we carry around with us into every interaction". Such knowledge is organised by the use of schemata, a set of cognitive shortcuts whereby a simple trigger activates all that we know and have experienced from previous encounters with a particular object, action or environment. The word "beach" for example brings with it expectations that may include yellow sand, blue sea, the call of gulls, children building sandcastles and adults playing beachball. Each schema provides a reference frame for interpreting incoming sensory data and presents a likely script (Schank & Abelson, 1977). Schemata and, in turn, the *umwelt* are continually updated depending on new encounters and experiences, so a visit to a volcanic island may create a revised schema that includes the possibility of black sand. Pinchbeck and Stevens (2005) suggest that levels of presence are affected by the interaction between sensory data and a user's schemata. In particular, bottom-up information that is incongruent with a schema may delay processing and interrupt presence. Familiarity only with pebbly beaches in East Anglia may make mention of sandcastles and beachball surprising. Conversely, this may also explain why Nunez & Blake (2006) found that, for flight simulator games, first-hand knowledge of flying actually reduced presence rather than enhanced it, as users were more aware of flaws in the content within the VE.

For a low-immersive environment, namely books, Gysbers et al. (2004) showed that a minimal description that provided less specific information evoked a greater sense of presence than a text abundant in detail. The author, Tracy Chevalier (see Minshull, 2011) illustrates this in a discussion of Guy de Maupassant's short story "The Necklace". Although the necklace is central to the plot, Chevalier points out that de Maupassant limits his description to just 4 words: "a superb diamond necklace". This allows the reader to imagine the necklace for themselves. It also allows that image to change over time. As Chevalier puts it:

> "As a teenager I'd pictured the necklace as a big net of diamonds that would cover most of your chest – something Elizabeth Taylor would wear, the sort of thing you need an impressive décolletage to display it on. Now, 30 years later when I imagine Mathilde's necklace, it is a simple strand of diamonds, very understated and elegant - Grace Kelly rather than Liz Taylor."

De Rijk et al. (1999) suggest that complex detail leads to sensory overload, destroying presence through fatigue. It seems reasonable, too, to argue that the more details provided, the more likely we are to find inconsistencies with our own experience. We may also be anxious that we are failing to construct an identical image to that which the author describes. Jones (2007) argues that the only details necessary are those that support the flow of action in the mediated world. Flow, in turn, facilitates the illusion of reality by minimising the salience of missing information.

As a schema, shaped by an individual's own experience, influences what they expect in terms of incoming information, it may also influence the allocation of their attention to reinforce those expectations. Achtley et al. (2012) showed that depressed individuals detected fewer words portraying people in a positive light (e.g. "winner") compared both with positive words that were unrelated to people (e.g. "sunshine") and with negative words. Participants with no history of depression showed no differences between word types. Such findings support one further integral element of the mental model of presence suggested by Schubert et al. (2001): that irrelevant information is suppressed.

2.2.3 Modal aspects of presence

Presence may not be experienced to the same degree in all modalities at the same time (Slater, Usoh & Steed, 1994). Tactile and kinaesthetic information about the "real" environment may cease to be attended to, as visual information from the VE triggers corresponding associations in non-visual modalities via the user's imagination. For example, in an experiment by Biocca, Kim and Choi (2001) participants manipulating a virtual hand-held tool reported a haptic response (the feel of its spring mechanism) despite receiving only visual information. The strength of this illusion correlated with levels of presence.

Presence is widely believed to be positively associated with a preference for the visual domain (e.g. Chen, 2000). Only recently has the contribution of other modalities begun to be explored. Jurnet, Beciu and Maldonado (2005) found no connection between presence and a preference for the visual mode when using a stimulus with important dialogue. Västfjäll (2003) observed that realistic aural rendering of events in mediated environments has come to be recognised as increasingly important in many multi-modal applications. This is borne out in a study by Skalski and Whitbred (2010) in which sound quality in video games had a greater influence than image quality on levels of presence. Similarly, enhanced auditory detail is more important than enhanced visual detail when it comes to users perceiving virtually-presented materials as “real” (Bonneel et al., 2009).

2.3 Mediated environments and visual impairment

2.3.1 Background to Audio Description

One group of people for whom the quality of the image is largely irrelevant is the visually impaired population. As mentioned above, media can be made accessible through the addition of a verbal commentary: Audio Description (AD). To the author’s knowledge, while there have been some attempts to create audio and tactile computer games for B&PS people (Delic & Sedlar, 2010; Sepchat et al., 2008) to date there has been no application of AD to mainstream VEs. Interactive AD at some level does exist in other contexts. Descriptive audio guides used at museums and galleries include audio instruction such as orientation information, providing directions to, and around the venue (e.g. http://www.vocaleyes.co.uk/) or verbally guiding a blind person’s fingers around a raised, tactile image (Canning & Fryer, 2010). In live AD contexts such as theatre, “touch tours” allow AD users to visit the stage before a performance, where they may handle props and costumes as these items are described (Szalwinska, 2009). Description is not limited to qualities, but includes affordances, and blind people are encouraged to operate, for example, a bell that will be used in the play (Pesky People, n.d.) or shown not only how to manipulate a lever, but also its effect in making a puppet horse's ears twitch (Fryer, 2010a). However, the most common application of AD is for recorded, screen-based content such as film and TV. Screen AD offers no opportunity for the AD user to interact with the describer. Even in the theatre where AD is delivered live during the play, such opportunities are limited.

Although the exact demographic of the AD audience is unknown, Ofcom (2009) suggests that, for TV, around one in five people with a visual impairment in the UK has used audio description (21%), and this figure is slightly higher (38%) for those with severe/profound visual impairment. There is no data on whether such users have congenital or acquired sight loss, but Ofcom's Code on Television Access Services confirms that "most potential users of audio description will have some sight, or will have had sight at some stage" (2008, point A2.23). Early studies (Pettitt, Sharpe & Cooper, 1996; Schmeidler & Kirchner, 2001) demonstrated cognitive and social benefits for AD users, although some (e.g. Peli & Fine, 1996) were conducted with sighted rather than B&PS individuals.

AD itself can be regarded as a form of mediation. By adding a verbal commentary, woven around the existing dialogue, to capture the visual elements of a scene, AD creates what Piety (2004, p.455) calls an "audio amalgam". In this way, an AV medium is translated into pure audio, so that for the AD user, a film or play comes to resemble an audio drama (Fryer, 2010b). This shift might not matter, except that, as we have seen, media form can affect presence (Freeman, Avons, Pearson, IJsselsteijn & Davidoff, 1999; Lombard & Ditton, 1997). Steuer (1992) developed a categorisation system, based on *interactivity* and *vividness*. With its limited *interactivity*, screen AD is akin to books, newspapers, film or broadcast TV. Live AD can be more interactive, akin to a Skype call or email. Steuer divides *vividness* into sensory breadth i.e. the number of modalities the media form engages, and sensory depth i.e. the degree of resolution of each modality. AD is a spoken commentary delivered via a single sensory channel: sound. Its method of delivery, via infra-red or radio headphones in the theatre/cinema or mixed into the soundtrack for TV/DVD, will to some extent affect its sensory depth, but it is proposed here that, for the mostpart, AD resides in the middle of Steuer's vividness spectrum alongside radio (Fig. 2.2).

The parameters governing the technology of AD delivery are outside the scope of this thesis. Instead the effect of AD on media content has prompted the following studies. AD content guidelines have been developed in a number of countries including the UK, Spain and Germany (see Rai, Greening & Petre, 2010 for a review). These guidelines are often contradictory and have been compiled with little research to inform them (Gerber, 2007). In the UK, provision of AD is recognised under the Equality Act (2010) as a "reasonable adjustment" for service providers of other AV media such as cinemas and theatres, but there are currently no guidelines for AD content in these settings.

Although AD content is driven by the visual information in the source material, there is time to describe *explicitly* only a tiny proportion of what is available *implicitly* to the eye. As with other media content, AD differs in style in the amount of detail included; the way the words are structured and, beyond that, delivered in terms of tone of voice, pace, pitch etc. All these factors may affect the way the information is received by the user, and, in turn, their sense of presence.

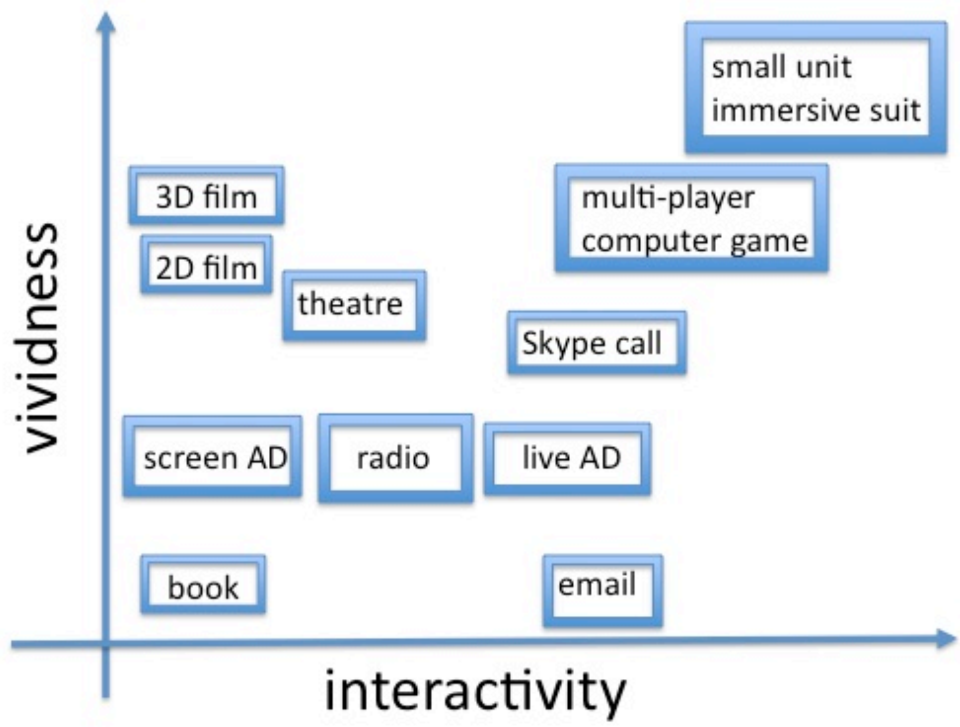


Fig. 2.2 Positioning AD (screen AD and live AD) alongside other media forms in terms of vividness and interactivity (after Steuer, 1993)

2.3.2 AV media: The sighted audience

In low-immersion, mediated environments direct perceptual information comes from a reduced number of modalities compared with the real world. In AV media, with which this thesis is mostly concerned, people with good sight and hearing enjoy a stream of congruent information from two modalities: vision and audition. Watching a bottle falling, for example, they also hear it shatter when it hits the ground. The two information streams support and complement each other, replicating the kind of environmental synchrony of auditory and visual stimuli that Bertelsen and de Gelder term “valid co-occurrence” (2004). As argued above, it is proposed here that such bimodal perceptual information activates crossmodal and semantic associations shaped by the experience of the viewer, filling in absent sensory data, to create a sense of “being there” (Fig. 2.3).

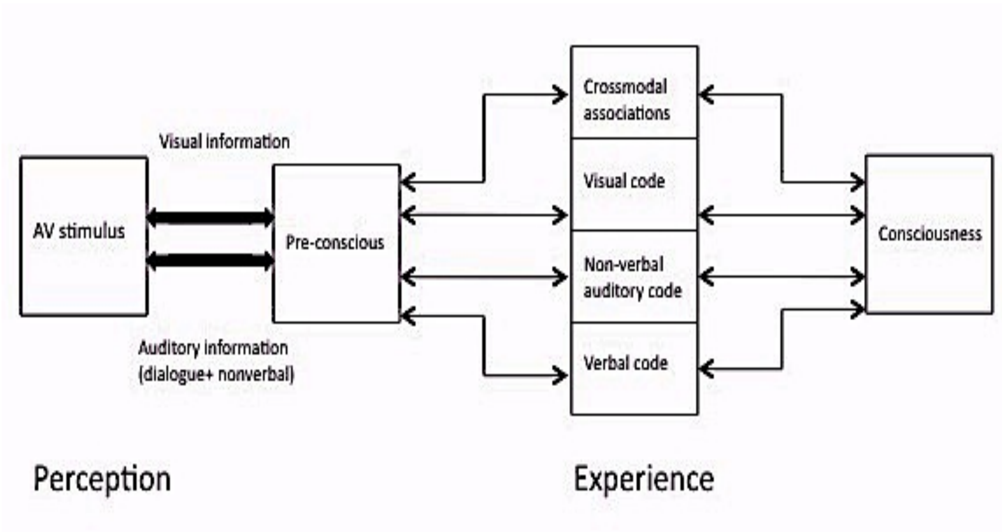


Fig. 2.3 Model of an AV user with no sensory impairment

Occasionally audio and visual streams will be deliberately incongruent to achieve a particular effect. For example, Tarantino’s film *Inglourious Basterds* is set in France in World War II, yet the soundtrack incorporates music by Ennio Morricone that is commonly associated with Westerns. This deliberate anachronism is at once a source of humour and hints at parallels between “goodies and baddies”, typical of a Western, that will be applied to 1940s France. The presence of non-diegetic music (i.e. music not arising from the action or environment where the scene is set) has implications for many of the dimensions of presence discussed above. It has been shown, in sighted people, to affect Ecological validity but not Spatial presence (Nunez, 2007) which is maintained through the visuals. When watching Hitchcock’s *Psycho*, for example, even though it is not natural to hear sharp, shrill violin shrieks as you are taking a shower, the mediated environment (the bathroom) is maintained by what you see.

2.3.3 The blind and partially sighted audience

It is possible that, for the blind audience, without access to the visual information stream, non-diegetic music may lead to a temporary "Break in Presence" (BiP: Slater, 1999; 2002) as the mediated environment ceases to persist for the duration of the scene. Similar BiPs may result when a non-speaking character affects the action, or when the translated words of a character who speaks in a foreign language are subtitled rather than dubbed, resulting in an inability to follow the narrative flow. More subtle nuances, as when the facial expression of a character belies their speech, may initiate a false schema in the mind of the blind user, affecting their interpretation of the drama. These issues will be explored further in Chapters 3 and 8.

In addition to information of which the sighted viewer is made consciously aware, the B&PS audience is also excluded from endogenous attentional biases induced by vision. Lingering shots that denote the start and end of a scene clarify event segmentation, which aids distribution of cognitive resources (Zacks, 2010) by indicating when one schema is no longer relevant and should be replaced by another. It could be argued that in the theatre this is achieved by lighting, fading to black at the end of one scene and brightening at the start of the next, or, within a single scene, visually refocusing the attention of the sighted audience from one part of the stage to another. When viewing static scenes, sighted people do not all fixate on the same part of the picture at the same time (Mannan, Ruddock &

Wooding, 1997). However, Smith (2006) has shown that the opposite is true for more complex, dynamic scenes. This effect, called “attentional synchrony”, even extends to sighed viewers mirroring a speaking actor who blinks, by blinking themselves after a 250-200 ms delay, and has been shown to correlate with neural processes including emotion. The link between affordance and attention outlined in Chapter 1 would suggest that in a play or a film, a sharp knife lying casually on a work surface, for example, primes sighted viewers to expect cookery or violence depending on the plot. Other visually-apprehended qualities which may or may not be consciously registered may include the length of the blade, the keenness of the edge, and whether the knife is gleaming or rusty.

In addition to, or perhaps because of missing out on such cues, the motivation of the blind audience to engage with an AV medium may be lower, if they have had previous experience of struggling to follow a programme or film.

2.3.4 A model of Audio Description

In order to address these problems, the visual stream in an AV stimulus is replaced by the verbal commentary. This identifies characters, and describes locations and actions. AD is predicated on the assumption that a verbal commentary can successfully replace visual information. This process is modelled in Figs. 2.4 – 2.6. For PS, the unreliable visual element of bimodal information coming from the source material will trigger intermittent crossmodal associations similar to those experienced by sighted people. Elements that are hard to see directly will be augmented by the verbal AD. For LB, bimodal associations stemming from earlier AV experience may be triggered directly by the auditory information from the soundtrack and indirectly by the AD, prompting an imaginal experience through a residual visual code. Crossmodal associations may differ, depending on the strength of that code, in line with age at onset of visual impairment. For CB, with access neither to the visual stream of information nor to a visual code, crossmodal associations will be triggered by auditory information alone i.e. the soundtrack and the AD.

The model provokes a number of questions. How do sight characteristics affect crossmodal associations triggered by vision and/or sound? Can words trigger those associations as effectively? Is visual experience necessary for verbal replication of visual information to have an effect? Does additional mediation (the verbal AD) increase cognitive load? Crucially does AD enhance or diminish presence?

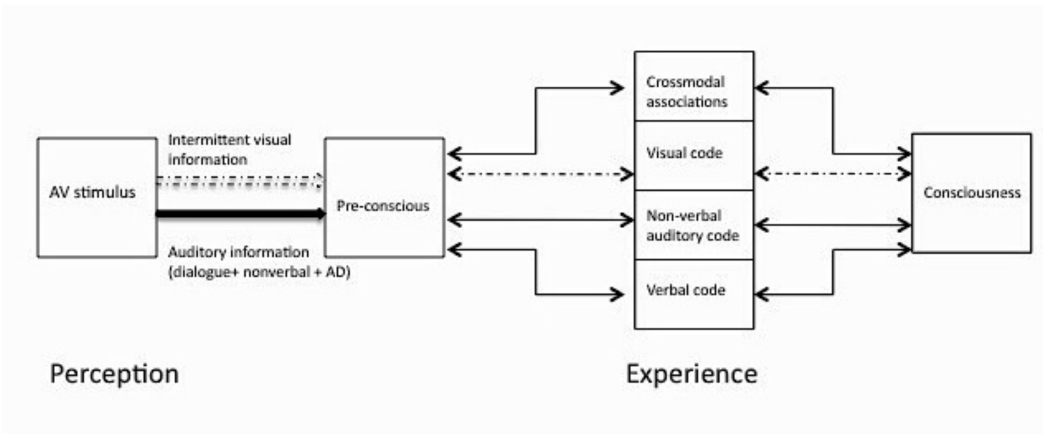


Fig. 2.4 Model of an AV user with partial sight (PS) watching[3] with AD.

[3] As B&PS people themselves use the term "watch" to describe the activity of engaging with AV media, that usage is respected in this thesis.

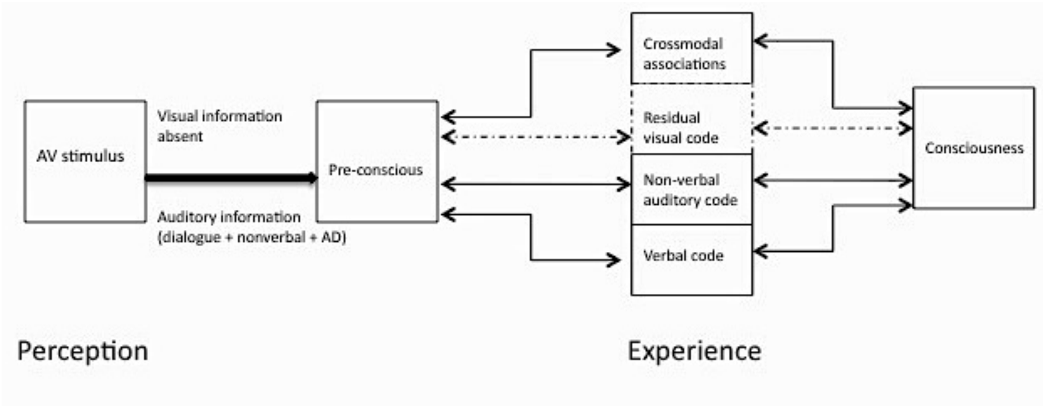


Fig. 2.5 Model of an AV user with late-onset blindness (LB) watching with AD.

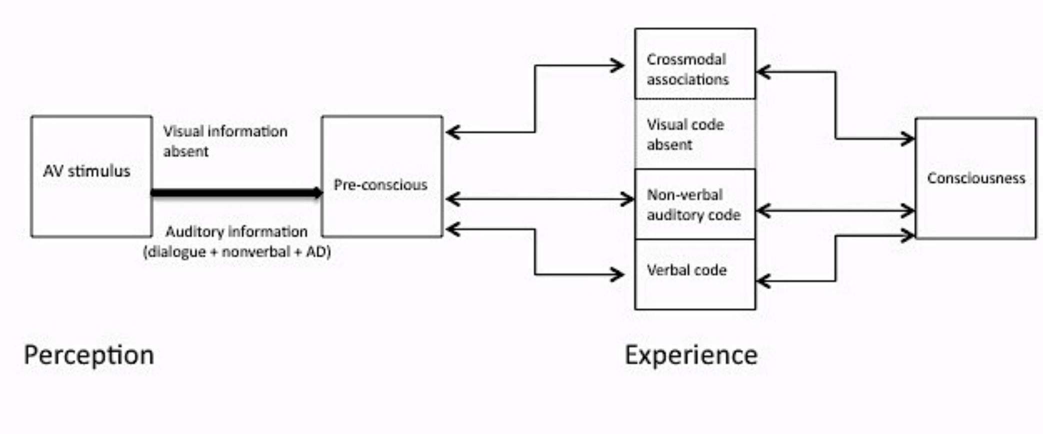


Fig. 2.6 Model of an AV user with congenital blindness (CB) watching with AD.

Given the hypotheses already discussed, there are a number of possibilities:

i: Sensory compensation: If improved auditory ability compensates for lack of vision, presence will be unaffected by missing visual information. Presence will therefore not be enhanced by AD.

ii: Linguistic compensation: If language compensates for lack of vision, presence will be enhanced by AD.

iii: Visual experience: If early visual experience is necessary for bimodal gains, visual information (conveyed verbally) will benefit LB/PS but not CB/EB. AD will therefore lead to greater levels of presence in the former than the latter.

2.3.5 AD and cognitive load

It has been argued that AD affects the medium *and* the message: both by changing an AV medium into pure audio and by translating visual information verbally. The latter may be more or less effective depending on the time available to squeeze AD in between bursts of dialogue and on the skill of the describer. Mousavi et al. (1995) suggest that strain on WM can be reduced when information is presented in more than one modality. For B people, with access only to the auditory modality, cognitive load may be greater. This may, in turn, reduce presence. However a number of studies have shown that B&PS out-perform S in tests of long- and, particularly, short-term memory such as digit-span tests (e.g. Hull & Mason, 1995). Rokem and Ahissar (2009) put this down to better auditory discrimination, which reduces load on WM by allowing B&PS to pick out speech more easily from background sound. However they also suggest that improved sensory coding of speech may allow B&PS to "chunk" verbal information, allowing it to be processed more effectively. This may be the result of specific training techniques, or experience, which in turn may lead to cortical changes. This idea is supported by evidence that the occipital cortex is recruited for memory tasks by B&PS but not S (Amedi et al., 2003).

Increased WM capacity in B&PS may override the danger of AD increasing cognitive load. An alternative possibility is that, by improving access to the content, AD may actually lower cognitive load, as AD users no longer have to struggle to piece together a whole story from incomplete parts of the narrative. In this case, presence may be enhanced by AD. However, there is also the factor of individual differences, in experience and ability, as well as sight characteristics and brain plasticity that might have an influence on load, and hence on presence. Arguably, visual terms, such as colour, and references to distant objects such as rainbows or outer space, may have more meaning (and therefore be easier to process) for those who have visual experience than those who do not. This may lead to a differential impact of AD content on presence, depending on sight characteristics.

2.4 Methods and measures

2.4.1 Methodology

2.4.1.1 Research participants

In order to understand visual cognition through the prism of blindness, many studies limit their samples to congenitally, totally blind participants. However, as more is uncovered about the visual process, the concept of "no sight" becomes problematic. Total cortical blindness caused by tumours or lesions such as scotoma may give rise to different cortical reorganisation compared to that caused by pathology of the peripheral sensory system (e.g. the congenital condition of anophthalmia where the eyes fail to develop). Cattaneo and Vecchi go so far as to suggest it is possible to "see" without eyes (2011) as the rest of the visual network remains in tact. Similar ideas have been put forward to explain "blindsight" (the ability of some blind people to recognise objects they cannot directly perceive): namely that, despite lesions, some islands of functioning cortex remain (Frendrich et al., 1992). Given the experimental benefits outlined above and the applied nature of this research, the experiments reported in the following Chapters have been carried out with participants from across the spectrum of visual impairment. This sample is more ecologically valid, reflecting the demographics of the blind population and, in particular, the audience for audio description. On a practical level, such a sample is also easier to find. All the visually impaired people who took part in the studies presented here are registered in the UK as either blind or partially sighted.

2.4.1.2 Measure of visual impairment

This thesis concerns visual impairment at the cognitive level. Instead of a clinical sight assessment, therefore, the following studies used a measure developed by Douglas, Corcoran and Pavey (2006) for the Network 1000 project. The project created a sample of 1000 people from the UK's B&PS population to provide data for a range of research questions. It used a 7 point scale based on the participants' self-reported level of functional vision, judged by response to the following:

> Which of these best describes your sight with glasses or contact lenses if you normally use them?
>
> 1: I have no light perception
>
> 2: I can tell by the light where the windows are
>
> 3: I can see the shapes of furniture in the room
>
> I can recognise a friend by sight alone if….
>
> 4: I'm close to their face

5: I’m at arms’ length away

6: I’m on the other side of the room

7: I’m on the other side of the street

For the purposes of this thesis, participants scoring 1 (i.e. agreeing only with the first statement) were regarded as being totally blind (no useable vision). Those scoring 2 or 3 were regarded as having a severe visual impairment (some useable vision). Those scoring 4 or above were regarded as partially sighted (considerable useable vision) even if they were legally registered blind.

A few participants had difficulty answering the Network 1000 questions, raising the objection that, for example, they could see shapes of furniture in some lighting conditions better than others, or on some days better than others depending on their general health or other stresses or concerns. Similarly, they might recognise a friend across the street if that friend was wearing a familiar red shirt and if it was a bright day, but not if the friend was wearing a new outfit, a drab outfit or the sky was overcast. For this reason, participants were also asked to decide themselves whether they had no, some, or considerable useable vision. These categories have been used in other research related to audio description (e.g. Schmeidler & Kirchner, 2001).

For the purposes of data analysis in some experiments, participants were further subdivided into the following categories: Congenitally Blind (CB) i.e. totally blind from birth; Early Blind (EB) i.e. totally blind from infancy (by the age of 3) such that they have no visual memory; Late Blind (LB) i.e. totally blind from the age of 4 years or above; Partially Sighted (PS) i.e. with some or considerable useable vision.

2.4.2 Measures

2.4.2.1 Qualitative Measures

Qualitative research is generally descriptive and, by its nature, not easy to quantify. It can include behavioural observation, field interviews, open-ended questionnaires, focus groups and one-to-one structured or semi-structured interviews. It has been used in presence research in a number of studies (e.g. McGreevy, 1993; Jacobson, 2001; Freeman & Avons, 2000). The advantage is that in-depth information is acquired in the user's own words, reflecting multidimensional responses to real-life experiences, rather than experiments tailored to the laboratory. In the research presented here, one-to-one semi-structured interviews designed to find out about the visually impaired audience's experience of audio description, and how that may link to presence, are reported in Chapter 3. This data is used illustratively to establish emerging themes. In Chapter 7 qualitative data is used more systematically to explore Spatial presence through the subjective experience of real-world navigation in S and B&PS. In other chapters participant comments are used alongside quantitative data to illuminate points of analysis. The disadvantage is that qualitative techniques generate a large quantity of data to be filtered making it vulnerable to subjective interpretation by the researcher. Individual responses may be hard to generalise. For that reason, this thesis employs both qualitative and quantitative methods.

2.4.2.2 Quantitative Measures: Questionnaires

The most common measure of presence is the post-test questionnaire. These have the advantage of making the data easy to analyse by quantitative methods, allowing comparison between different groups of participants, so a subjective experience can be generalised to a wider population. Questionnaires can also assess a number of dimensions of presence simultaneously. The disadvantage is that questions are necessarily answered in hindsight and therefore prone to biases of recall. They also rely on the participant understanding the intention of the question with no opportunity for clarification. Some questionnaires are relatively untested and may not be appropriate for all types of media or all types of participant.

For the presence experiments in this thesis (Chapters 6 & 8) the presence measure selected was the ITC-SOPI (Lessiter et al., 2001). This has been extensively used in presence research over the past decade. The ITC-SOPI asks participants to rate 44 statements on a 5-point Likert scale (1 = *strongly disagree*, 2 = *disagree*, 3 = *neither agree nor disagree*, 4 = *agree*, 5 = *strongly agree*). The ITC-SOPI assesses 4 dimensions of Presence: sense of physical space or Spatial presence (e.g. "I felt characters or objects could almost touch me"); Engagement (e.g. "I felt involved in the scene"); Ecological validity (e.g. "the scenes seemed natural") and Negative effects (e.g. "I felt I had a headache"). The dimensions are analysed separately as they are not necessarily equally affected by different determinants of presence. Given a possible range of scores from 1 – 5, an average greater than 2.5 is the threshold at which at least some degree of presence is attained (J. Freeman, Nov 2013, personal communication).

2.4.2.3 Physiological measures

Recently, attempts have been made to find more objective presence measures using physiological responses such as heart rate (HR) and galvanic skin response (GSR) (e.g. Ward et al., 2002; Mandryk et al., 2005), also known as electrodermal activity (EDA). These are particularly useful for measuring the affective aspects of presence. They have the advantage of being continuous and can monitor fluctuations in presence as, for example, a film progresses. As they are time-coded, physiological data can be synchronised to events on screen. This allows the impact of specific moments to be analysed, rather than a one-off rating to cover the whole experience. The disadvantages include practicality, as such measures are limited to use in an experimental setting, and confounds may arise from individual differences, for example heart rate variability (HRV, see below) may be high either as a direct response to the stimulus or due to an individual's general ability to cope with stress. This makes it difficult to group results or generalise from one subject to another. Measures of HRV and EDA are used in Chapter 8.

2.4.2.3.1 Heart Rate

HR is controlled by the autonomic nervous system, subdivided into the excitatory sympathetic nervous system (SNS) and the inhibitory parasympathetic nervous system (PNS). For stress-related emotions such as fear, the SNS excites the heart's "pace-maker", the sinoatrial node, leading to an increased heart rate. This readies the body for flight or fight. In situations with no threat, the PNS takes over. Its inhibitory effect on the sinoatrial node lowers heart rate. The SNS and PNS operate antagonistically such that heart rate speeds up either from increased activity in the SNS or decreased inhibition from the PNS (Applehans & Luecken, 2006). Between them, they lead to continual fluctuations in interbeat intervals (IBIs) i.e. the length of time between heartbeats. Sympathetic activation takes a while to take effect, peaking after a 4s delay and taking 20s to return to baseline. By contrast parasympathetic inhibition is very rapid, peaking after 0.5s and taking only 1s to return to baseline. These autonomic processes are triggered by the Central Autonomic Network (CAN). This coordinates exogenous and endogenous stimuli, allowing an individual to adjust their physiological, behavioural and cognitive responses appropriately. HR is commonly measured by variability (HRV) namely the variance from normal interbeat intervals (i.e. Normal-to-Normal: NN) recorded over a specified duration.

2.4.2.3.2 Electrodermal activity

Another common physiological measure to assess user experience is EDA. Specific sweat glands (eccrine glands) on the palms of the hand and the soles of the feet respond to psychological triggers rather than simple temperature changes. These glands act as variable resistors such that the resistance of the gland decreases, whether or not sweat actually breaks out on the skin (Stern et al., 2001). This change in resistance can be measured with electrodes attached to two fingers of the same hand. EDA has been shown to correlate with arousal. Mandryk et al. (2005) showed that, when playing a computer game with a friend rather than against the computer, EDA increased in line with subjective ratings of Engagement. EDA was negatively correlated with frustration, but positively correlated with challenge and fun. EDA has also been used to evaluate responses to low-immersive media such as text and television (Jones, 2007) making it particularly suitable for the media forms used in the current research.

2.5 Aims and plan of thesis

The following chapters explore aspects of perception and experience in people with varying sight characteristics. Rather than moving along a single path from one study and its results to the next, it can be regarded as a search for congruency across different types of study. Petrie et al. (2006) have noted the benefit of using remote data collection techniques for participants with disabilities, given that such participants can be hard to find. Yet they also point out that for qualitative data, face-to-face encounters may be more desirable. Here, findings from experimental, descriptive, remote (online) and field-based research are integrated into a synthesised model of how visual impairment impacts on perception, real life experience and, by extension, engagement with the virtual world.

All the studies reported received ethical approval from Goldsmiths College, University of London. Chapter 3 presents subjective reports of presence in mediated environments through interviews with blind and partially sighted AD users. Chapter 4 tests so-called “universal” sound symbolism effects, to see whether reduced access to a visual code affects other bimodal (in this case auditory-haptic) associations. Chapter 5 addresses evidence for and against auditory and linguistic compensation in the imaginal domain, comparing the effect of sight characteristics on clarity of auditory imagery and vividness (breadth and depth) of imagery prompted by words or sound effects. Chapter 6 explores the impact of sound effects on presence in a dynamic, low-immersive medium (audio drama) comparing S and B&PS. This raises questions about Spatial presence and Ecological validity. These are addressed in Chapter 7 by investigating spatial mental models in S and B&PS when recalling navigation routes in the real world. Chapter 8 considers the role of emotion on presence and how it may be enhanced or reduced by the addition of AD. Chapter 9 discusses what may be learned from the preceding studies about cognitive processing in the visually impaired audience, its effect on presence and the links between presence and AD. It also considers directions for future research.

Chapter 3 Study 1: Experience of Audio Description: What do Blind and Partially Sighted Users Think?

3.1 Introduction

Ross (2003) claims that disabled audiences have seldom been the focus of academic study. Research with B&PS audiences, as a subset of the disabled population, is even more limited. As discussed in the previous Chapters, both bottom-up and top-down information processing will be affected by visual impairment. However, the impact of this on engagement with media content is not clear. Bertrand Russell (1917) distinguished between “knowledge by description” and “knowledge by acquaintance” arguing there is no way of going from one to the other. If this is the case, the question arises whether “knowledge by acquaintance” is sufficient to create a sense of “being there”? This will be tested experimentally in Chapter 8. Following discussions with an AD users’ focus group, Evans and Pearson (2009, p. 387) suggest “AD cannot successfully substitute for the images if the viewer has no visual referent to allow their imagination to follow the onscreen narration”. Yet, is a visual referent essential to presence? How do B&PS use the verbal information of the AD? Does it differ according to aetiology of visual impairment?

In order to gain a better understanding of the experience of the B&PS audience and to inform subsequent studies, semi-structured interviews, lasting up to 90 minutes, were conducted with a sample of AD users covering both genders and a range of ages and sight loss characteristics. Although the main focus of this thesis is on AD for screen-based media, respondents also commented on their experience of live AD e.g. at theatrical performances. The interviews resulted in a profile of the AD audience in the form of AD personas developed for the International Telecommunication Union (ITU) standards committee on access. This is presented in Appendix A.

3.2 Method

3.2.1 Participants

Participants were recruited through blind organisations e.g. the RNIB, the Macular Disease Society (MDS), VocalEyes, East Suffolk Association for the Blind and personal contacts. This resulted in the following sample: N = 16 (9 male), aged 31 – 74 years, m = 54.31, SD = 17.17. Sight status: CB = 2 (2 male); LB = 7 (4 male); PS = 7 (3 male). Participant demographics are shown in Table 3.1. All were regular users of AD. None were paid for their participation.

3.2.2 Materials

The following questions formed the basis for each interview:

Q1. When listening to AD, do you build up an image in your mind?
Q2. How important is it that your mental image mirrors what other people see?
Q3. What is your worst experience of AD?
Q4. What is your best experience of AD?
Q5. How would you like AD to develop in the future?

ID	Gender	Age	Sight status	Age at onset of impairment	Visual acuity*	Aetiology
001	M	73	LB	60	1	Retinitis Pigmentosa
002	M	71	LB	7	1	Accident
003	M	65	PS	6	5	Congenital cataracts
004	M	59	CB	Birth	1	Rubella
005	M	63	CB	Birth	1	Retinopathy of prematurity
006	M	37	PS	Birth	5	Juvenile macular degeneration and central vision loss
007	M	31	PS	20	4	Optic atrophy from brain tumour
008	M	32	LB	4	1	Persistent hyperplastic primary vitreous
009	M	54	LB	7	1	Retinitis Pigmentosa
010	F	74	LB	45	1	Retinitis Pigmentosa and glaucoma
011	F	73	LB	37	3	Myopia, detached retinas, cataract, glaucoma, dry macula degeneration
012	F	74	LB	5	1	Optic nerve damage
013	F	37	PS	Birth	5	Colaboma, micro-ophthalmia nystagmus and subsequent retinal damage
014	F	38	PS	27 registered blind, but onset from childhood	5	Lawrence Moon Bardett Biedl syndrome
015	F	31	PS	8	4	Retinitis Pigmentosa
016	F	57	PS	9	5	Meningioma of the optic nerve and optic atrophy

*Visual acuity coding (after Douglas et al., 2006). Which of these best describes your sight with glasses or contact lenses if you normally use them? 1 = I have no light perception; 2 = I can tell by the light where the windows are; 3 = I can see the shapes of furniture in the room. I can recognise a friend by sight alone if: 4 = I'm close to their face; 5 = I'm at arms' length away; 6 = I'm on the other side of the room; 7 = I'm on the other side of the street.

Table 3.1 Participant demographics

Most participants were interviewed individually, except for two married couples (ID002 & ID010; ID003 & ID012) where husband and wife were interviewed together. All interviews were recorded and transcribed by the researcher, except for ID009 who responded to the questions in writing.

3.3 Results: Quantitative Analysis

The interviews ranged from 1,059 – 14,029 words. Responses related to AD in theatre, cinema and television. Answers to Q1 and Q2 were analysed quantitatively as well as qualitatively. They were coded as follows: Q1 “Do you build up an image in your mind?” 0 = never; 1 = sometimes; 2 = yes; Q2 “How important is it that your mental image mirrors what other people see?” 0 = not important; 1 = quite important; 2 = very important. Results are shown in Figs. 3.1 and 3.2. Both CB participants reported never building up an image, LB were more or less evenly divided and PS participants showed a greater tendency towards building an image. Non-parametric tests, however, showed no significant effect of sight status: Q1 CB/LB: Mann-Whitney U = 3, $p = .195$; CB/PS: Mann-Whitney U = 1, $p = .064$; LB/PS: Mann-Whitney U = 17.5, $p = .343$. Results for Q2 were more equally distributed and again sight status had no significant effect: CB/LB: Mann-Whitney U = 6, $p = .748$; CB/PS: Mann-Whitney U = 4.5, $p = .434$; LB/PS: Mann-Whitney U = 19.5, $p = .493$. There were also no significant effects of gender for either question: Q1 M/F: Mann-Whitney U = 29, $p = .779$; Q2 M/F: Mann-Whitney U = 30.5, $p = .908$.

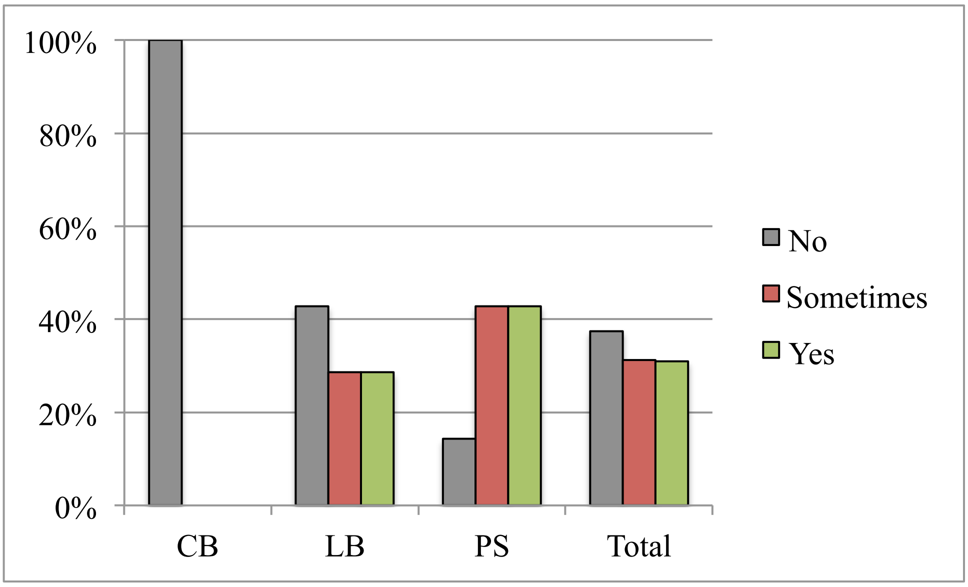


Fig 3.1 Percentage of respondents by sight status to Q1:

"Do you build up an image in your mind?"

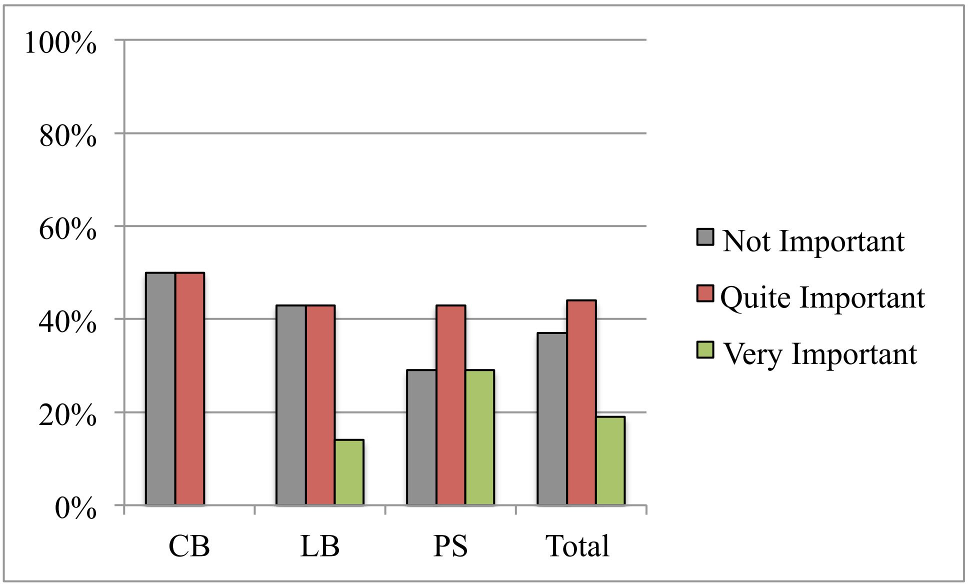


Fig. 3.2 Percentage of respondents by sight status to Q2:

"How important is it that your mental image mirrors what other people see?"

3.4 Qualitative analysis: Emergent themes

The interviews produced many spontaneous comments to suggest people with visual impairment can immerse themselves in mediated environments such that they have a sense of “being there”. One participant suggested that the experience of a blind person could even be enhanced compared with that of a sighted person, because, with no visual reminder of media form (such as wobbly scenery or the frame of the cinema or television screen) there was less need to suspend disbelief. However, most felt that presence was hindered by lack of access to visual information. Dialogue and non-verbal sound did not always enable a person with a visual impairment to follow the plot, let alone access the “full picture”. This led to a sense of exclusion or general disengagement. In general, comments suggest that the information given in the AD compensates by providing visual details in verbal form.

There was some suggestion that the type of verbal information required depended on visual experience. Participants with PS desired more information when the quality of the visual stream of the source material was reduced, for example by low levels of lighting or the use of black and white film rather than colour, or the AD user’s position in the auditorium. Other participants expressed a preference for more minimal rather than extensive AD but this was not consistent as regards aetiology. CB as well as LB and PS mentioned being “transported” by mediated content, even though they were aware that their mental representation may not directly reflect the action as seen by sighted people. All dimensions of presence were referred to (indirectly): Spatial presence, Ecological validity, Engagement and affect, as well as Social presence that came either through connecting with characters in the scenes or with fellow members of the audience. Comments also provided behavioural evidence of presence, with participants reporting that mediated content triggered involuntary physical responses such as wincing or flinching. It also became clear that the best experience of AD was when the user was unaware of it. Conversely poor AD could lead to Breaks in Presence. Positive and negative experiences of engaging with mediated environments are discussed in more detail below, illustrated by participants' comments.

3.4.1 Mental Imagery

As the quantitative analysis shows, participants varied as to whether they always, sometimes or never used the verbal information from the AD to build up a mental representation. ID009, who lost his sight at the age of 7, described it as "a sequential process of adding details as the AD provides them". ID008, who became blind at the age of 5, said:

> "I always build up a sort of picture of what I think the stage looks like, even if the stage bears absolutely no resemblance to the picture I have in my head. I think that's based on the stage I used to act on at school. So I always imagine a stage being a rectangular space… and then when the audio describer is describing actors moving around I have a picture in my head of them going left, right, forward, back… I'm quite lucky because I had sight until the age of 5, so I'm reasonably able to create pictures in my head, because I've got a sort of visual framework."

For those who did not visualise the scene, one reason was the speed of the stimulus. An older participant whose sight had slowly deteriorated said: “If somebody picked up that cup, I think I do imagine a hand going round the handle and picking it up, but it’s a flash, it’s very quick” (ID011). A younger, partially sighted participant, ID015, put it this way:

> “In films quite often the action moves quite quickly and…there’s not much time really to build up a picture, so I don’t really tend to build up pictures in films very much…. But … the other night I was watching some sort of nature programme and that was talking about turtles and they were burying things in the sand and stuff and I built up quite a picture then”.

Not all participants with visual experience constructed a mental image. Participant ID001 gradually lost his sight over many years and has had no light perception for the past ten years. He said:

> “When I go to an audio described [performance]…do I imagine it as a visual thing? No I don’t think I do but then I don’t imagine a visual world that much anyway except when somebody asks me to do it…by which I mean that I can see you sitting there but what I’m seeing is your voice…my seeing, in that sense, is making inferences from the sense data I’ve got, it’s not making visual images in my head…So if I don’t visualise it, what do I think is happening? I’m imagining action…I make inferences from the dialogue, from the sound of the actors … and the AD

> might well contribute, you know, "she rushes across and hits her", in the audio description…[but] I don't visualise it. What I do is I imagine it."

Neither of the two CB participants used descriptive information to build up a mental representation, but appreciated being told about visual details that were pertinent to the plot. As ID004 explained:

> "I picture not an image of the person, no, because that's a bit alien to me… it's more facts that tell me about the person, like grey hair, balding, moustached…Description of the rooftops wouldn't probably tell me a great deal because I don't know what rooftops look like, but descriptions of the room would, I'm used to that… so the impression of the room, untidy, 60s-looking, you know, furnished in 80s style, these things I would understand."

3.4.2 Embodied perception and first-person perspective

Contrary to Evans & Pearson's supposition, having a visual referent was not always regarded as necessary in order to appreciate the audio description. ID008 said: 'Whereas some blind people find it very hard to visualise things, I'm quite good at it, so I can visualise a forest without having been in one." ID012 explained:

> "I do like to know how people see things …because it broadens things for me. Although I lost my sight when I was 5 and probably don't have much

> actual [visual] memory, nevertheless I love to hear about it because then I can build up a picture in my own mind about it and it's enriching".

ID005 described imagining the unfamiliar by building on an experience he *was* familiar with. So if a description included a reference to a galloping horse, for example:

> "What I do, I put myself on that horse by feel. So I have been on a horse, I've felt a horse, a big one, a racehorse, I haven't been on a racehorse but I've been to Ascot … loads of times, so I know what the sound of 20 galloping horses sounds like, galloping 10 yards from me, it's amazing. So I would, yeah I would imagine myself feeling that horse. I wouldn't imagine it in any visual sense but I can imagine being in the saddle, I know what it's like holding the reins, so yes, I'm not saying what I would imagine would be correct but I would strive to imagine it, yes I would, in a feely, sound sense. A bit like dreaming."

It has been suggested that a first-person perspective increases a sense of immersion (McMahan, 2003). ID012 reported the use of this perspective, creating a sense of embodied presence through rich, non-visual imagery:

> "I don't have a visual picture in my mind and yet … when you said a clock ticking, well I imagined a room, my home and um…sitting in a particular armchair and I knew everything in that room, you know, where it was, what it was and the sensation, you know it's a cosy room, round

> table etc. etc. Now all of that came into my mind but I wasn't looking at it…but I was imagining myself in it and knowing what was in it… so it's not quite, not quite actually looking but it is quite a full, I don't know what you'd call it…".

The researcher suggested "mental representation" and ID012 agreed.

3.4.3 The impact of AD on dimensions and determinants of presence

The suggestion, critical to presence, that AD can bring with it the illusion of non-mediation was borne out by all the participants. ID012 summed it up this way:

> "When it works [AD] almost becomes part of the play, somehow. Because it's almost at a sort of subconscious level and it sort of flows as if …it's obviously not part of the play but it feels part of it."

ID015 confirmed AD improves the flow: "it makes it a bit more seamless and more enjoyable, a bit more sort of a smooth sort of experience."

3.4.3.1 Mediating media form

IJsselsteijn (2002) suggests that we develop media schemata that help us disconnect from, for example, horrific or gory scenes in VEs as we know they are not real. Media schemata are more likely to be activated by visual indications of media form, such as the edge of the TV or cinema screen, or the proscenium arch in a theatre. As B&PS have fewer cues to activate media schemata, they might be more likely to treat the content as though it were real. ID008 pointed this out, suggesting that, in a theatre production he, as a blind person, might find the settings more real than the sighted audience:

> "For a blind person, it's easier to imagine the shift from the palace bedroom to the forest because they can put it in their head. Whereas a sighted person has to try to forget the scenery, the stage, the hall…"

Interestingly, although AD was felt to work best when it was least noticeable, several participants mentioned the importance of verbal reminders of media form. ID007 used the example of a theatre production of *Chitty Chitty Bang Bang*:

> "[You're told] there's a car and you think so it's a car then [but] it could be just a cardboard cut-out which people see and I wouldn't...at those times, I feel cheated, not cheated, that's a little strong but … I would say it is important to know if it's a real car or just a car cut-out, a bloke standing in a cardboard car."

This may be related to Social presence (discussed below). However, ID001 enjoyed AD in the theatre, but not screen AD because:

> "The conventions of the theatre I think I've got in my head and they work for me with audio description – it's a different kind of audio description from what I've had in, admittedly the two films I've seen, and some television, and it's something to do with getting inside the medium. For television I don't think this matters quite so much because I can imagine

> the formats of television, but in film, I want to quote, unquote, "see" a film as a film, and not hear it as a badly narrated …er…radio play."

AD guidelines discourage the use of technical cinematic terms (see Fryer & Freeman, 2012a; 2012b) and a reminder of media form was not considered essential by all participants. ID007 recognised the limitations of AD when it came to describing camera angles and editing style:

> "I miss not having it but I understand the capabilities of the AD and the fact that they can't describe absolutely everything otherwise you wouldn't be able to hear the text."

3.4.3.2 Social presence

For those who expressed an appreciation of media form, this was at least partly linked to a desire to share the experience of the sighted audience. Research highlighting the importance of shared visual experience in a mediated environment originally focused on business and training applications (e.g. Buxton, 1992; Romano et al., 1998) and multi-user devices in online gaming (Kraut et al, 2002). Mantovani & Riva (1999) point out that advantages in terms of communication and collaboration derive from individuals sharing a common reference grid in any type of environment, be it virtual or real. More recently Social presence has been shown to correlate with increased satisfaction with the VE (Bulu, 2012). As for the experience of watching a mediated experience with one's peers, Shedlosky-Shoemaker et al. (2011) found that social influence can enhance both enjoyment and psychological "transportation".

B&PS participants were certainly eager to share in a common experience both within the virtual environment presented on stage or on screen and in the real environment i.e. the theatre or cinema auditorium. ID008 said:

> "I want to try and get as close as I can to what a sighted person would experience when they go to the theatre, so I want to be able to experience [the actors'] actions and the movements and the looks and all the things that, as a blind person, I don't get. The best way of putting it is listening to other people laugh and not knowing why they're laughing. Because...it's quite an isolating experience to kind of hear other people laughing and to be thinking what's going on, why are you laughing? It might be that if I knew, I might not find it funny but it's just knowing that something is happening that's having an effect on everybody else but I'm out of that experience."

ID013, who has partial sight, agreed that AD could help with this:

> "In Comedy of Errors there was so much going on, that while I could see the characters moving around the stage, I wasn't always entirely sure of what they were doing, but the AD told me, so I was roaring with laughter at the same time as everybody else."

Although participants varied in the degree to which they wanted their mental representations to mirror what the sighted audience saw, most (10/16) agreed that it was quite or very important, as the impact could be felt not just during the event but afterwards too. As ID009 put it:

> "I am a firm believer that blind people suffer from visual illiteracy – that is, when sighted people gather and discuss a movie, TV programme, or even a media image, they share the knowledge of what the thing looks like, with all the data points that image knowledge contains. I, as a blind man, lack this essential data, so I am at a disadvantage in this discussion."

ID015 also felt there was a social element involved. She said:

> "I think the more iconic an image is for the way it looks, the more it matters to me that my picture is accurate. For example most educated people around the world I suppose, or certainly the western world, would have an instant picture in their minds if you said "Big Ben" to them. So I would like an accurate picture in my mind - everybody knows what it looks like, or at least there's probably a common perception of how it looks…maybe in reality close up it's grubby or smaller than perceived, or bigger, or crumbling or something? So I want to share in the common perception."

The remaining 6 participants felt that constructing the same mental image was neither important nor in fact possible. ID005 described his mental representations as rich and multi-sensory but argued they could not be similar to what sighted people saw as they were without a visual dimension. He said:

> "I can imagine being in a field and hearing the sound and feeling the breeze and smelling the grass but there'd be nothing visual. There'd be no flowers, there'd be no sky, there'd be no clouds, there'd be no shapes, do you know what I mean? There'd just be a general feel of open air. There might be a smell of grass, animals, there might be sounds of birds, sheep… in other words I can only imagine what my store of human experience can tell me."

Not only visual experience but also individual differences in perception play a role here. ID015 pointed out that sighted people do not all see the same thing. She used the example of colour:

> "I have PAs (personal assistants) that help me get dressed and stuff and like I'll ask them what tops are in the wardrobe at the moment, and everyone sees colours differently and they all describe them differently and I think it's quite personal so… it doesn't bother me that much and I think it's hard to know whether you have got the same image 'cos everyone sees things differently."

For one congenitally blind man (ID004), the principal advantage of AD was that it gave him a propositional understanding of the sighted world:

> "It's quite amazing because actually I learn things about people generally, about their gestures because I don't know what gestures are like, or looks…I learn a lot about how people behave, how the sighted world behaves and how it works in terms of visualisation."

3.4.3.3 Spatial presence

In terms of Spatial presence, ID002 explained his experience of AD in this way:

> "Oh crikey, well if they're describing movement and stuff you're following the movement aren't you? I mean if the girl lights a cigarette and leans back and picks up a cup – you're following that aren't you pet? You're in that living room with the story."

However, several participants pointed out that, for this to be achieved, the balance between AD and the sound from the source material needs to be right. As ID004 put it:

> "I think every time the describer speaks you're being, I mean the focus is being shifted…in some part away from the action. I know it's a difficult one because people say well you're still in the room and you're having it described but you don't get the ambient sound of the room when the describer's talking because [the volume of the soundtrack is] lowered."

ID015 agreed:

> "Sometimes in plays you can really get sucked into the atmosphere you know and the feelings that are being built up, and sometimes another person speaking [i.e. the audio describer] can distract from that and I prefer to sacrifice a bit of information for being immersed in the atmosphere more."

3.4.3.4 Ecological validity

For people with partial sight or a visual memory, AD can aid Ecological validity by, for example, adding to information about appearance. ID006 said:

> "On Sunday we had a film on for children – Dennis the Menace – and had the AD and it was good like because obviously I hadn't seen it before, and didn't know who the kids were or who anyone was…and [the describer] said Dennis is in the red and black top, the stripy top and the other bloke's in the brown scruffy coat so you have a rough idea what it looks like, you know...cos it adds colour to it, doesn't it? You're not just thinking of a scruffy old man, you know roughly what he looks like."

The importance of visual cues, such as facial expression, was acknowledged by all participants although it could also be a source of confusion. As ID008 put it:

> "I'm always intrigued about when they describe people looking at each other in a certain way. Cos I often find that kind of very hard to

> understand, so they often say you know "he looks at such and such crossly" or "he has a thoughtful expression". The first time I heard that I thought, "how do you know he's not constipated!" [laugh] So they are useful things to have, but...I find them harder to understand. And that might be because I don't have the visual experience."

For one congenitally blind participant (ID004) facial expression was more useful than other aspects of appearance:

> "I guess knowing looks and knowing gestures would be for me more important than descriptions of the people because I think that's…often more important to the plot. If someone scowls or looks happy that's more important than whether or not they're wearing a red dress or blue shirt."

The fact that some information about appearance added little to his mental image did not reduce Ecological validity as it was part of his everyday experience. He said:

> "I don't think I have a great need to be the same as others in the sense that I want to know exactly what other people are seeing because, well I guess because I'm not used to that. I guess that's where it's coming from."

3.4.3.5 Engagement, affect and empathy

Many studies have shown that emotions elicited by media content can affect engagement and presence (see Riva et al., 2007). Mood in media is maintained by stimuli in both the auditory and visual modalities, such as music, sound effects, facial expression and lighting. While these might not advance the plot, it has been suggested that they help activate schemata in the minds of the audience, setting up empathetic behavioural responses such that the audience reacts as if the mediated events are real. Behavioural responses can also be prompted by the combination of the AD and the soundtrack, as ID015 explained for the film 127 Hours (2010):

> "It's the one where the guy goes climbing and he falls and gets trapped and ends up having to cut his arm off. It was brilliant because had it not been for AD that film would have been completely inaccessible to me because most of the time there was only him, so there was no dialogue...but like, when he was cutting through his arm, there was music and there was him cutting through his arm, and there was the audio describer and I thought they all worked perfectly together and I had a real sensation of [sharp intake of breath] ooooh, this is really nasty and horrible and I thought the sound that they used when they actually got to his nerve, it was like a kind of twanging painful sort of sound and, like, the describer, and him screaming, they all worked really, really well together…I mean it was almost like the description was part of the work, like as a whole and that was really fantastic."

For another partially sighted participant [ID006], too much graphic detail could be a disadvantage:

> "I watch Casualty. That's got [AD] on. And … obviously I hate watching it if I'm eating my tea, my dinner… so I close my eyes but obviously [the describer will] tell you then there's blood gushing out – oooooh [jokey scream] And so I'm sitting there, closing my eyes and [the describer's] telling me like 'he's just cracked into her scalp' or something like that and there's blood oozing out and like ooooh, I'll have to turn the sound off as well."

3.4.3.6 Attention

Recent experiments with sighted people using eye-tracking have shown that AD can act as an exogenous cue in guiding attention. Vilaró & Orero (2013), for example, compared eye-gaze for a short sequence from the DVD of Danny Boyle's film *Millions* (2004). The DVD menu begins with the yellow cockerel mobile of the Pathé Logo, its shadow projected on a white wall. Once the film settings have been selected, and the user presses "play", the cockerel appears again, modified such that the shadow now features a halo. Only the group watching with AD reported noticing the difference, and a heat map showed the eye-gaze of those viewers had shifted up to focus on the halo itself. Similarly, Krejtz et al. (2012) showed that sighted children watching 2 clips from an educational animated programme with AD fixed their gaze on items of interest significantly more often and for longer than children watching without AD. Comments from partially sighted participants in the present study reinforced the evidence that AD helped them direct their residual vision. As ID016 commented for theatre AD:

> "You like to know what side [the action is] going on or whether it's in the middle or which side of somebody somebody is, or…all that sort of thing."

However, those without sight reported achieving the same effect by sound:

> "If you're sitting close enough you can get that sense of people talking from a distance or people talking from one side of the stage to the other." (ID008)

For those with partial sight, lighting and contrast can influence when AD becomes important:

> "If the lights are on one person and they're really well lit up and they're in brightly coloured costumes or whatever I could probably see that fine, but you might not see, you see, in the background in a darker area there's someone doing something they shouldn't be."

Arguably, AD should reflect the experience of the sighted audience, who may also fail to notice what is going on in the background, but several participants enjoyed the fact that "sometimes we know more what's going on than the sighted people" (ID016). One participant, now in his 70s, who lost his sight as a child summed it up like this: "I would say [AD] gives you depth of vision" (ID002).

3.4.3.7 Breaks in presence

Good AD was equated to an experience of non-mediation. For ID001:

> "If I come out of the play having really, really enjoyed the play and not having been acutely conscious that I'm blind then the AD has worked. And if someone then says 'you know what was it about the AD that was so good?' it is a meaningful answer for me to say well I actually didn't notice it."

Conversely, experience of poor AD centred on elements that disrupted presence, by making the user aware of the describer. ID007 said:

> "I think the only [AD performances] I can remember properly are the ones where it didn't work because I can describe them in detail, but the rest are all much of a muchness really."

All participants pointed out that too much detail in AD could contribute to excessive cognitive load. As ID015 put it:

> "Sometimes, although the description is giving extra information which is interesting, it gets kind of tiring listening to it, and yeah, it makes it more hard work."

The addition of redundant information was also a source of irritation. ID14 said:

> "The ones that make me laugh are 'There's a knock at the door' and you hear 'bang bang bang' and you think that was quite obvious. Maybe who's at the door would be nice to know but not that there's going to be a knock at the door."

ID001 expressed it this way:

> ID001: "Just occasionally in the AD in the theatre I do find stuff that's superfluous because given the sort of build up in a dramatic scene you can more or less tell what's likely to happen and then if somebody tells me there's a really angry expression on his face…"
>
> LF: "…you can hear it…"
>
> ID001: "Of course you can bloody hear it! I mean I spend all my life doing it and maybe it's a slight underestimation of the skills acquired by people with sight loss over time. Because they learn to interpret the world or the stage in terms of its sound and the movements they can hear and they're making inferences from those things all the time, you know, and they're actually pretty good at it."

However, omission was also a common complaint. ID014 commented on an audio described performance of the musical *We Will Rock You*:

> "One part during the show they're singing a song and it's got all these iconic pop stars who died – ended up with Freddie Mercury of course – but [the describer] didn't tell you who they were. At one point there's a statue that rises up from the lake and she just went 'a statue has arisen' well you guess it's Freddie Mercury but it would have been nice to have that, you know, cos it might not have been, and just key things like that were missing."

ID002 felt it was important for the describer "not to skirt round stuff like sex or blood." This not only excluded him from content but from the shared audience experience. As he put it:

> ID002: "Sometimes you get a glossing over feeling."
> LF: "What leads you to suspect they're glossing over?"
> ID002: "Because it's bound to be worse than it is."
> LF: "Maybe your imagination makes it …"
> ID002: "Your neighbour knows but you don't know - that's the difficulty."

ID009 also felt he sometimes received "a Cliff Notes experience". In addition, he found it frustrating when AD failed to update visual information. This relates to Pinchbeck and Stevens' (2005) suggestion, discussed in Chapter 2, that incoming information incongruent with a schema can delay processing and therefore cause a BiP:

> "I recently watched a movie in which the female character spent most of the time barefoot. There came a scene… in which she kicks a package across a prairie. Throughout this scene… I imagined her barefoot and wondered how she was able to kick this package for what seemed miles with her bare foot. Only at the end of the scene did the narrator describe the character as wearing short buckskin boots. This late detail caused a glitch in the flow of my viewing."

For partially sighted participants, flow was not only affected by poor AD, but also by style factors, relating to the source content (see Jones, 2007, discussed in Chapter 2). ID013, who has partial sight, does not always feel the need to use AD, however:

> "In action films that have very fast sequences I might have to say to someone sitting next to me, you know, what happened there? Um… because I don't always see the fast action."

Similarly, she finds problems with:

> "Anything in black and white. Because the contrast disappears. And I often use a strategy to identify my characters by remembering what they're wearing and what colour their clothes are, or I remember the colour of their hair or something like that and I know it's the same person. And in black and white movies you can't do that. And so I remember

> when I went to see *Schindler's List*[4] just being totally lost because to me it looked like a whole series of disconnected frames um…with no continuity, none of the characters seemed to be the same, it was just, completely disconnected."

ID007 agreed that AD became more important for

> "Older black and white films, which are a little fuzzy round the edges… I watched an old war film – I say watched – in fact I didn't understand any of it…it was too fast paced and I'm colour blind as well so black and white, well it's sort of grey, grey and grey all the same colours to me, so I [didn't] understand it at all unfortunately."

3.4.4 AD and the linguistic compensation hypothesis

In line with earlier studies, participants were enthusiastic about the benefits of adding a verbal commentary to access AV media. When asked what they wanted of AD in the future, most participants simply replied "more". ID015 specified increasing AD for different types of art forms such as dance. Interestingly, ID007 wanted AD to mediate the real world as well as the virtual world. He said:

> "If I made a lot of money I'd have an audio describer with me at all times, telling me facial expressions or at least following me from a distance and I could have a chip [implanted], so they wouldn't have to say in front of me – that person doesn't get your joke, or whatever…they're walking

[4] Spielberg's 1993 movie that begins and ends in colour, but for most of the time is in black and white

away, you know, or they're trying to steal money or whatever."

Given that, in Gibson's ecological theory of visual perception, "to see things is to see how to get about among them" (1986, p.232), this suggests that verbal information is a useful substitute for those who cannot see for themselves. This supports the linguistic compensation hypothesis proposed in Chapter 2 suggesting that words as much as sensory input from non-visual modalities compensates for visual deprivation.

3.5 Summary of Study 1

B&PS experience AV media not through bimodal channels but by relying on sound alone. This reduction of sensory breadth might be expected to reduce presence. AD replaces the missing visual information with words i.e. semantic information. This might seem a poor substitute yet these B&PS participants spontaneously mentioned experiences of presence, and cited dimensions of presence as ways to distinguish good from poor AD. From their comments it seems that, at its best, AD aids presence by easing cognitive load, smoothing flow and increasing a sense of Spatial presence, Ecological validity, Engagement and affective and emotional connection with the mediated characters. It can also enhance Social presence not only with on-screen characters but also with fellow members of the audience. By contrast, AD at its worst exhibits elements associated with BiPs including omission or overly dense and misleading information that, by increasing cognitive load, breaks the flow. This suggests that AD has as great an influence on presence by changing media content as by changing media form.

Redundant verbal information was criticised partly because it underestimated the ability of blind people to identify information from non-visual sources and partly because the additional commentary destroyed presence by interrupting ambient sound. Both these aspects would support the theory of auditory compensation. Yet linguistic compensation was also evident. This raises the question of whether there is a qualitative difference between non-verbal and verbal auditory information that may reflect differences in processing with and without sight? Furthermore how may that affect presence? The effect of words and non-verbal sounds are compared in Chapters 5 and 6. The next Chapter addresses the impact of visual impairment on sound symbolism testing crossmodal associations between sound and touch.

Chapter 4 Study 2: Visual Impairment and Sound Symbolism: The Bouba-Kiki Effect

4.1 Summary

Chapter 3 gave an insight into the impact of visual impairment on perception and presence. The areas where AD was felt to be lacking, such as omitting details that were salient or supplying ones that were redundant, illustrate both what information is lost in the absence of sight and what information can be picked up through other senses. The degree to which B&PS participants actively formed a mental image from the words of the AD varied, but perception and experience clearly both play an important role in shaping that imagery. The model of AD in Chapter 2 raised the question of whether non-visual crossmodal associations might be different for those with no or reduced access to a visual code. The following experiment draws on a familiar paradigm to test the role of visual experience in bimodal associations between two non-visual modalities: sound and touch.

4.2 Introduction

Since the 1920s, researchers (studying sighted populations) have demonstrated a correspondence between words and shapes. In experiments by Köhler (1929, 1947) English-speaking adults consistently matched the nonsense words "Maluma" and "Takete" with outline images of a rounded shape and a jagged, star-like shape respectively. Since then the same effect has been demonstrated across a range of populations, including English- and Swahili-speaking school children presented with the nonsense words “Uloomo” and “Takete” (Davis, 1961) and adults given word pairs from their non-native language (Koriat & Levy, 1979). Ramachandran and Hubbard (2001) ascribe the 95% correspondence rate that has been observed between image-shape and word-choice for “Bouba” and “Kiki” to a form of synesthesia. They suggest that the “Bouba-Kiki” effect results from contiguous areas of the brain processing the visual outline of the shape (rounded/star-shape) and the rounded or angular appearance of the speaker’s lips when enunciating the vowels. They also proposed a link between the sound contours of a word and a shape’s visual appearance.

The “Bouba-Kiki” effect is not limited to bimodal links between sound and vision. Word associations have also been demonstrated in the modality of taste (e.g. Deroy & Valentin, 2011; Crisinel et al., 2012). Gallace et al. (2012), for example, showed that crisps are deemed more “Takete” than a soft cheese, while chocolate is rated more “Kiki” if it is mint-flavoured. The researchers argued that these word-food associations stemmed from differences in flavour rather than shape or texture, so the “Bouba-Kiki” effect may be independent of vision.

In addition to the appearance of the rounded/spiky shapes another visual influence is the orthographic form of the words. The rounded outline of the letters B and O in “Bouba”, for example, might encourage an association with the rounded shape, while with the spiky forms of the letters K and I in “Kiki” would do the same for the jagged shape. To this end, Maurer, Pathman and Mondlock (2006) asked pre-lexical toddlers to associate four pairs of rounded/spiky shapes (3 pairs of 2D drawings/cut-out shapes and one pair of 3D objects modelled in clay) with four pairs of contrasted nonsense words differing in their vowel sound. Although overall the toddlers' bias was not as strong as that of a control group of adults, the youngsters still associated the rounded forms more consistently with rounded vowel sounds, and angular forms with non-rounded vowel sounds than would have been expected by chance.

Nielsen and Rendall (2011) questioned whether vowels alone were responsible for the "Bouba-Kiki" effect. They argued that most studies contain a fundamental flaw, using word-pairs where the vowels in one word resembled one of the shapes more closely than the other. To avoid this orthographic confound, they devised a number of studies swapping the vowel/consonant relationships of previously-used word-pairs (e.g. takete/maluma became takouta/malimi) and devising new ones using consonants with an incongruent orthographic form (e.g. M has a spiky appearance but a sonorant sound). When the words were presented graphically, 80% of participants made the expected associations based on the consonants, while their performance for associations based on vowels was 51% i.e. around chance level. Nielsen & Rendall also found that mode of presentation (visual/aural) affected the relative influence of vowels/consonants. When the words were presented aurally, the strength of the association for both vowels and consonants was reduced. For consonants, only 58% of participants made the "expected" associations. For vowels, the "Bouba-Kiki" effect disappeared completely, with only 42% of participants mapping the words to the expected shapes.

Given the different impact of vowels and consonants, Nielsen & Rendall's findings suggest sound is more important than orthographic appearance in creating the "Bouba-Kiki" effect. They argue that this reflects a difference in the auditory quality of consonants. Animal observation studies (e.g. Rendall, Owren & Ryan, 2009) have shown that primates and other species emit shrill and staccato (strident) sounds at moments of high arousal and aggression, and smoother, legato (sonorant) sounds in positive social situations such as grooming or foraging. It is possible that the "Bouba-Kiki" effect is strongly linked to an auditory pattern of spectral density and attack whereby strident vocalisations are matched to negative, and sonorant vocalisations to positive patterns of affective and social behaviour. More recently, Bremner et al. (2012) have demonstrated sound-shape symbolism amongst the Himba people of Northern Namibia who have no written language, again demonstrating the importance of sound.

As the auditory element alone, however, cannot explain the difference in results between pre-lexical toddlers and adults it seems likely that some balance of visual-auditory correspondence underlies the "Bouba-Kiki" effect. Although Gallace et al. (2012) argue that the word-food associations, stemming from differences in flavour rather than shape or texture, show the effect is independent of vision, their participants were presented with a visual (rather than auditory) word stimulus and arguably particular tastes may have triggered visual imagery that mediated participants' choices. A mint leaf, for example, has a jagged outline as well as a sharp flavour.

As mentioned earlier, testing visual-flavour matches with the Himba, Bremner et al. (2012) had surprising results. Namely that the Himba did not map spiky shapes to carbonated water and rounded shapes to still water; nor did they map bitter chocolate to a spiky shape but preferred to match it with a rounded shape in the opposite direction to Westerners (Ngo et al., 2011). While the chocolate-shape choices could be explained by the sounds of the Himba words used to denote bitter and sweet, Bremner and his colleagues put the water-shape mapping down to visual conditioning. In the West, for example, brands of carbonated drinks often feature angular motifs (e.g. Spence, 2012; Spence & Gallace, 2011a). As suggested in Chapter 1, cultural conditioning may not only be determined geographically. Within Western culture, CB, EB and even LB individuals are likely to be unfamiliar or at least less familiar with the appearance of letters and advertising graphics than people who are sighted.

To explore the impact of visual impairment on crossmodal associations, this study presented haptic equivalents (2D and 3D models) of Köhler's outline drawings to B, PS and S people. Participants were allowed to feel but not look at the object pairs and asked to decide which was "Kiki" or "Bouba". If there is a direct bimodal correspondence between sound and touch there should be a similarity of response between those with sight and those with a visual impairment. However, if access to a visual code (through visual experience) affects non-visual associations then differences between sighted people and those with a visual impairment should emerge. A difference in response would also be predicted between LB/PS who have visual experience and CB/EB who do not.

4.3 Method

4.3.1 Participants

An opportunity sample was drawn from staff and visitors to the offices of the Royal National Institute for Blind People (RNIB) in London and from personal contacts. All those who took part did so on a voluntary basis, with no payment for their participation. This resulted in 122 participants, categorized as follows:

S: N = 80 (41 male), mean age 42.88 years, range 20 – 82 years;

B&PS[5]: N = 42 (24 male), mean age 48.9 years, range 24 – 80 years (CB = 6; EB = 0; LB = 17; PS = 19). One PS participant had congenital cataracts and therefore had no early visual experience. For more details see Appendix B1.

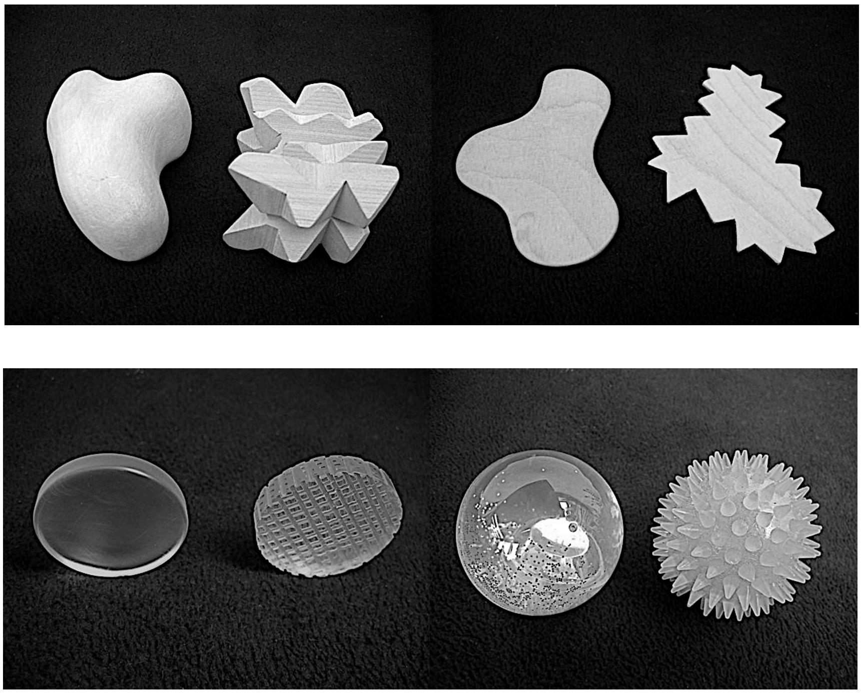
4.3.2 Stimuli

The stimuli comprised 4 pairs of shapes (see Fig. 4.1). Pairs A – C were specifically made for the experiment. The objects for Pair D were bought commercially. Pairs A and B were made from wood and designed to mimic Köhler's line drawings as closely as possible; Pair A in 3 dimensions (3D), Pair B in outline. Pairs C and D were made of synthetic materials. The discs of Pair C were identical in shape, but differed in texture. The spheres of Pair D were consistently spiky/smooth all over. Each pair of objects was presented in a black cotton bag, measuring 250mm x 250mm, fastened by a drawstring.

[5] For the purposes of this study, participants with severe/profound sight loss (i.e. little or no light perception) have been termed 'blind'. Participants with moderate/mild sight loss have been termed 'partially sighted'. This includes participants with moderate sight loss who are registered blind.

4.3.3 Procedure

The researcher followed the script shown in Appendix B2. The bags were handed over one at a time. The participant was asked not to look inside the bag, but to feel inside, and to bring out either Kiki or Bouba. The sequence of bags was counterbalanced across the sample. Half the participants were asked to bring out Kiki from the first and third bags, and Bouba from the second and fourth bags. For the remaining participants, this order was reversed. After selecting from all four bags, participants were asked to give a reason for their choices.



A B

C D

Fig. 4.1 Bouba and Kiki: Shape Stimuli

Pair A: 3D models made from wood (jelutong) each measuring approx 100 x 70 x 60 mms. One is bulbous and smooth in all dimensions, the other irregular and spiky. Pair B: 2D plywood cut-outs, 4mm thick, 115mm x 70mm at their widest points. One has a smooth outline, the other spiky. Pair C: cast acrylic discs, 40mm in diameter x 7mm thick. Both have the same outline, but a different texture. The surface of one is smooth, with chamfered edges and the other has been cross-hatched all over so it feels rough. Pair D: commercially available plastic balls – each about the size of a tennis ball (approx. 650mm in diameter). One is smooth, the other covered in rubbery plastic spikes.

4.3.4 Scoring

The presentation of each bag counted as a separate trial, making 4 trials in total. Participants scored 1 per trial if the object they took from the bag matched the expected word (e.g. the rounded object was identified as Bouba). If they took the incongruent object from the bag (e.g. rounded object identified as Kiki) participants scored 0. Thus the total score, summed across the 4 trials, ranged from 0 - 4. A total score of 2 would indicate performance at the level of chance.

4.4 Results

4.4.1 Quantitative analysis

The S group showed a robust “Bouba-Kiki” effect in the haptic-auditory modalities, mapping the shape to the expected word on 3.59 out of 4 trials. The B&PS group mapped the shape with the expected word on 2.57 out of 4 trials. While this was significantly higher than chance ($t(41) = 2.09$, $p = .043$), a one-way ANOVA showed it to be significantly lower than the score for the S group: $F(1,120) = 15.68$, $p < .001$. The shapes chosen as “Bouba” on each trial are shown in Fig. 4.2. Superimposed are the numbers of participants in each sight group who mapped “Bouba” to either the rounded or the spiky shape.

4.4.1.1 All participants

Of the S participants, 84% (67/80) chose as expected for all 4 pairs, selecting rounded objects as Bouba and spiky objects as Kiki; only 5% (4/80) consistently chose in the opposite direction. The remaining 11% (9/80) were inconsistent, choosing a spiky shape sometimes as Kiki and sometimes as Bouba.

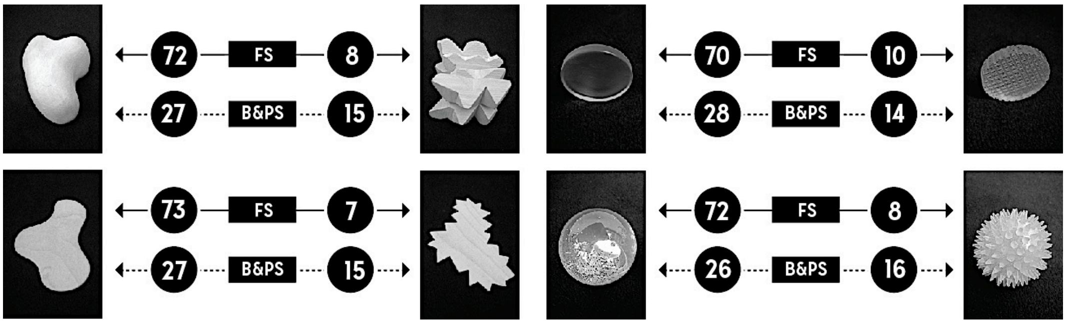


Fig. 4.2 'I would like you to bring out "Bouba" for me.' Number of participants choosing a shape as "Bouba" for each of the 4 object pairs Full Sight (FS) N = 80; Blind and Partial Sight (B&PS) N = 42.

By contrast, only 55% (23/42) of the B&PS participants chose in the expected direction for all 4 pairs; 26% (11/42) consistently chose the opposite, and 19% (8/42) were inconsistent. Those with impaired vision showed a significantly different pattern of mapping words to shapes from their sighted peers: χ^2 (2, N =122) = 14.40, p = .001.

4.4.1.2 B&PS participants: No visual experience versus visual experience

6 of the B&PS sample were congenitally blind. This group performed at the level of chance: 50% (3/6) mapped words to shapes in the expected direction and 50% (3/6) in the opposite direction. Planned comparisons with Bonferroni corrections showed that this was significantly different from the S group (mean difference = 1.59, p = .001) but not the LB/PS group (N = 36; p = .733). Group means for each trial are shown in Fig. 4.3 however these results should be interpreted with caution because of the large variation of group size. In particular, the CB group showed wide SD. The score for the LB/PS group showed no correlation with the percentage of life they had been visually impaired (r = -.177, p = .301) nor with their degree of visual acuity (r = -.168, p = .328).

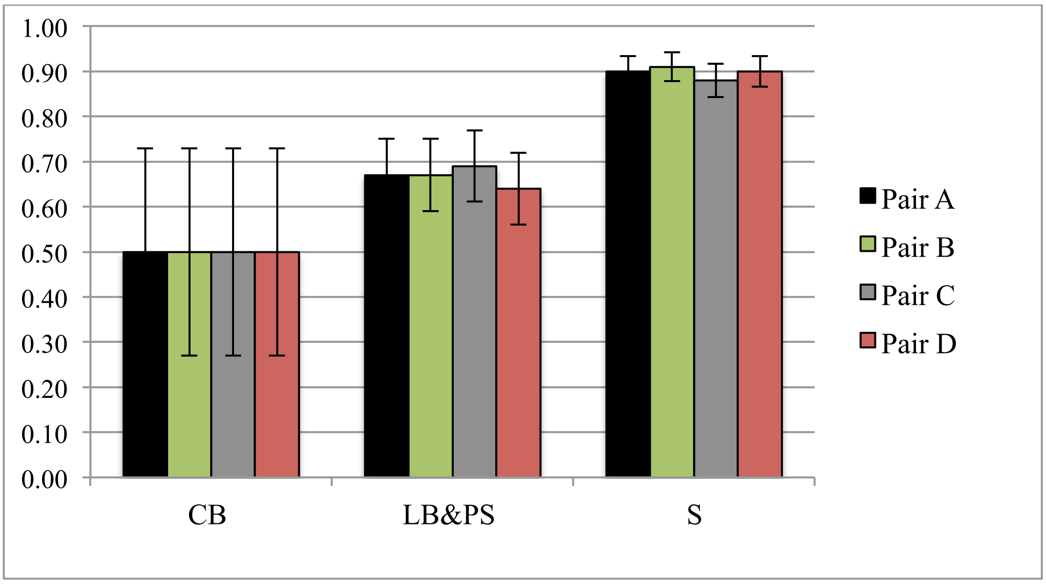


Fig. 4.3 Group means (SEM) per trial (max = 1)

4.4.2 Qualitative analysis

81/122 participants gave a reason for their choice (Fig. 4.4). Of the 58 S participants who did so, 60% (35/58) cited the sound of the word (e.g. “Kiki sounds jaggedy; Bouba sounds smooth”); 17% (10/58) cited orthographical appearance (“I was picturing a K in my head”); 8% (5/58) reported associating Kiki or Bouba with a name and their object choice was determined by perceived gender differences (e.g. “Kiki is a female name so I chose the rounded shape as Kiki”); 14% (8/58) gave other reasons for their choice, such as “It was the first object I picked up”. Of the B&PS participants who gave a reason, 61% (14/23) cited sound; 4% (1/23) a name and 35% (8/23) gave other reasons (“it was a pure guess”). None of the B&PS group cited the look of the letters, even if their sight problems had developed late in life.

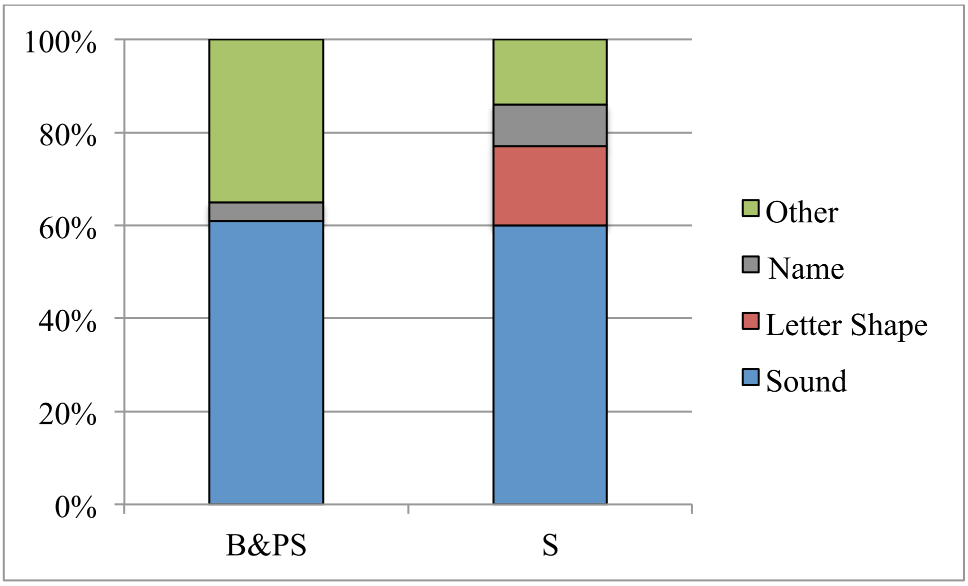


Fig. 4.4 Reasons given for object choices (by sight group)

4.5 Discussion

The “Bouba-Kiki” effect has been demonstrated in the visual-auditory modalities for over 80 years. It has been shown to override linguistic and cultural boundaries, and does not appear to be dependent on familiarity with particular letter shapes. It has even been extended to include taste. This study shows that, for sighted people, the “Bouba-Kiki” effect also embraces the haptic-auditory modalities. Aurally presented words and objects that could be touched but not seen created a robust “Bouba-Kiki” effect. People with a visual impairment who had residual vision or had had some visual experience also associated sharp/jagged shapes with Kiki and rounded/smooth shapes with Bouba at a level higher than chance, although the effect was significantly less strong than in their fully sighted peers. The small number of people in this study who had no visual experience (CB) chose at the level of chance (50%).

Sighted participants were able to access a visual clue in the shape of the researcher's lips/mouth shape while enunciating the words. However, the lack of consistent direction from those with congenital or profound sight loss would suggest that what Ramachandran and Hubbard term the "synaesthetic correspondence between visual appearance and vocalisations" (2001, p. 29) is not by itself sufficient to explain the "Bouba-Kiki" effect. Implicit visual associations are more likely to have played a role.

This study revealed a range of strategies for choosing which object was Kiki and which Bouba. The look of the letters was important for 17% (10/58) of the fully sighted participants who gave a reason for their choice: "I was visualising a K"; "a K has prongs"; "B is rounded". Regardless of how recently participants had developed sight problems, none with a visual impairment cited the look of the graphemes as influential.

This lack of connection with orthographic appearance may be enough to explain the difference in performance between sight groups. Amedi et al. (2001) showed that both visual and tactile recognition of objects activates a part of the object-responsive cortex in the lateral occipital complex (LOtv) where, more recently, bimodal visuo-haptic neurons have been identified (Tal et al., 2009). Lacey et al. (2009) in a review of studies on visuo-haptic convergence argue that the LOtv is supramodal, and can be driven by geometric shape information regardless of the modality it is acquired through i.e. vision or touch. However, a study by Holtby and d'Angiulli (2012) with S participants showed that identification of haptic

pictures decreased in the presence of visual interference. This suggests that haptic stimuli are encoded in the memory, at least in part, via a visual code.
It is possible that those with visual experience who cited an auditory connection ("Bouba's reminiscent of 'baby' and fitted the rounder shape", "Kiki sounds like 'kinky' – all nooks and crooks") may have also been drawing on visual imagery. Each word may have triggered a mental image, so participants were visualising, for example, the appearance of a baby or a kinked object, and then comparing that image with the unseen object in their grasp.

In the audio-visual domain, Zangenehpour and Zatorre (2009) found that, for sighted people, even brief habituation to visual and auditory stimuli that were presented simultaneously led to a response in the primary visual cortex being automatically triggered when participants were subsequently exposed just to the auditory stimulus. The current study suggests the visuo-haptic modalities are similarly tied. Interestingly in Zangenehpour & Zatorre's study the reverse situation did not hold: being exposed to the visual element of the bimodal stimulus did not trigger a response in the auditory cortex. Just as visual dominance is evident in perceptual AV and visuo-tactile cues (such as the rubber hand illusion discussed in Chapter 1), it is possible that vision retains its dominance when triggered indirectly through imagery. Arguably the small minority of sighted people who failed to demonstrate the "Bouba-Kiki" effect may have been individuals with below-average ability to evoke vivid visual images i.e. verbalisers rather than visualisers (Paivio, 1977). This would be interesting to test in future research.

The majority of participants (60%) blind or sighted, who gave a reason for their choice, reported being influenced by the sound of the words. Supporting Nielsen and Rendall's theory that not only vowels contribute to the "Bouba-Kiki" effect, 3 participants specifically mentioned the vowels ("in Bouba the vowel sounds are round") and another 3 picked out the consonants ("Kiki – the consonants sound sharper"). Most participants simply cited the word as a whole ("Kiki sounded spikier, more angular"; "Bouba is softer and smoother"). 7% (6/81), regardless of visual acuity, linked Kiki or Bouba to a name – sometimes resulting in unexpected choices: "Bouba's like Pumba[6] so I chose the chunky one"; "Kiki is a female name and female equals rounded". However, this was not consistent: "Kiki is the awful singer who used to screech and sounds hard and angular". Sound associations amongst participants with sight loss were sometimes unexpected. To a man who became blind as a child, Kiki "sounded more stream-lined" so he linked it with the smoother shapes. A congenitally blind man who also chose in the unexpected direction said "Kiki sounds more like an animal, more organic; Bouba sounds more artificial and textured". Another simply said, "Kiki sounds softer".

Some participants with impaired sight reported making a haptic association, comparing the feel of the object with one that felt similar, either in texture or outline. A blind man who felt the 2D plywood objects first, reported that the one with the rounded outline should be "Bouba" because the shape "resembled a boomerang". Another associated the textured disc with the rough skin of a kiwi fruit and made a link between "kiwi" and "Kiki". It is possible that the word

[6] Pumba is a warthog character in the film The Lion King (Disney, 1994)

association was made first, and the haptic connection was secondary. However some associated words sounded less similar. A man who lost his sight before the age of 5 reported that the jagged 2D object "felt like three Christmas trees" and Kiki "sounded more spiky". Struiksma et al. (2011) point out that touch is the primary way in which blind people identify objects in the world around them. It may be, then, that they are more likely to make a concrete (object-to-object) rather than an abstract (shape-to-shape) association.

Interestingly, Oberman & Ramachandran (2008) found that children on the autistic spectrum performed poorly on the standard visual/auditory version of the Bouba/Kiki test, mapping the word to the expected shape only 56% of the time. Autistic-like traits exhibited by some (but not all) children with congenital blindness include echolalia, pronoun reversals and formulaic speech (Hobson & Bishop, 2003; Pring, 2004). However, Perez-Pereira and Conti-Ramsden (2004) point out that autism is a genetically-based, neurodevelopmental disorder and therefore quite different from blindness, which is a peripheral sensory impairment. Furthermore, autistic-style traits are only found in children who are totally congenitally blind. Pring (2002) suggests that although blind children may have delayed speech development, the majority catch up. Of the 6 CB adults in the current study, 5 had a college degree. It is unlikely that the B&PS participants overall showed any higher incidence of autism than the general population.

Bremner et al. (2012) argue that the "Bouba-Kiki effect" in the visual-auditory modalities appears to be universal, driven by phylogenetic factors, while other forms of crossmodal shape symbolism (namely shape-flavour matches) are culture-specific. This study suggests that, even within the same culture, sighted individuals can pick up on regularities in their environment that are not as easily accessed by those for whom visual information is restricted. It would appear that visual impairment limits the strength of the “Bouba-Kiki” effect even when stimuli are presented in non-visual modalities.

4.6 Summary of Study 2

Study 2 suggests that crossmodal associations, in this case auditory/haptic, are influenced by experience and perception. Although for the B&PS group there was no direct correlation between total score and visual acuity nor the length of time they had lived with a visual impairment, still those with visual experience exhibited the Bouba-Kiki effect (albeit significantly less strongly than their fully sighted peers) whereas those without visual experience did not.

Although Kiki and Bouba are auditory stimuli, in the sense that they are nonsense words, they are nonetheless verbal rather than non-verbal sounds. For some B&PS participants the sound of the word was important (whether it “sounded” smooth or spiky), for others, the meaning i.e. its semantic association was the critical factor. This was the case for those who associated “Kiki” with a name. As discussed in Chapter 1, whether language is amodal (not linked to any particular modality); multimodal (leading to simulated perceptual experience) or supramodal (a combination of the two) is still under debate. If it is amodal or supramodal, then it is likely that the nature of mental imagery would vary when triggered by words (that have semantic content) compared with imagery triggered by non-verbal (modal) sound. This is of particular interest given the concerns, raised in Study 1, that passages of verbal commentary (AD) obscure auditory information from music or sound effects and may consequently reduce Spatial presence or Ecological validity. Study 3 explores differences in imagery arising from verbal stimuli (descriptions of sounds) and non-verbal sound stimuli (sound effects) between people with full sight and those with a visual impairment.

Chapter 5 Study 3: A Comparison of Mental Imagery Prompted by Word Cues and Non-verbal Sounds

5.1 Introduction

AD seeks to communicate to its audience, in words, visual information they are unable to access directly. As we have seen, there is some debate as to what information AD needs to convey. Study 1 showed that AD users could be irritated by redundant description that repeats information audible through sound effects. "He walks along a gravel path", for example, might be obvious from the ambient sound of footsteps crunching at a measured pace. Peli and Fine (1996) found that for a nature documentary, some AD utterances conveyed details that could be picked up from the standard soundtrack. Their research was with sighted participants and they proposed that blind people might be able to glean even more, being "presumably more skillful at gaining information from nonvisual sources" (p.379). As the volume of the soundtrack is often reduced for each passage of description, Lopez and Pauletto (2009) go further, suggesting that by obscuring non-speech sound, AD limits the use of the listener's imagination. The assumption of both studies is that, compared with sighted people, those with impaired vision have a heightened response to non-verbal sound.

This is in line with the auditory compensation hypothesis. As discussed in Chapter 1, blind people have been shown to have a compensatory advantage in some aspects of auditory perception. To what extent that covers all aspects, or transfers to the ability to imagine sounds, is much less clear. For example, Noordzij, Zuidhoek & Postma (2007) explicitly compared visual, spatial and auditory imagery in B&PS and S. They found that when asked to imagine sets of three sounds (e.g. "knocking on the door", "banging a nail" & "firecracker") B&PS participants were no more successful than S in picking the odd-one-out.

5.1.1 Auditory versus multisensory imagery

Auditory imagery, defined as "a perceptual-like experience of an auditory stimulus in the absence of that stimulus" (Hubbard, 2010, p.324) has attracted much less experimentation than imagery in other modalities, in people both with and without sight (see Hubbard, 2010, for a review). This is, at least in part, due to the difficulty in demonstrating, either by behavioural or by neurophysiological methods (e.g. Yoo, Lee & Choi, 2001), that participants are generating imagery only in the desired modality. As Study 2 demonstrated, visual imagery, for example, may be unwittingly triggered by haptic or auditory inputs. It may be that in response to sound cues in Noordzij, Zuidhoek & Postma's study, S were benefitting from access to visual imagery, and thus able to compensate for B people's supposed auditory advantage.

Although Struiksma et al. (2009) suggest that spatial imagery is unique in not being tied to a specific input modality (see Chapter 1) the multisensory nature of real-life experience argues in favour of this being true of imagery in general. Research for the marketing industry has shown that perception of taste, for instance, is influenced subconsciously by the shape and design of the packaging (e.g. Gallace et al., 2012). It was proposed in Chapter 1 that the modalities contributing to a multisensory mental image would be weighted according to their importance, with vision dominating in sighted people. The question arises, therefore, whether sound dominates multisensory imagery in those without access to visual memory. If not, do other non-visual modalities provide some form of compensation or do those without visual experience simply create less vivid imagery than those who have, or have had, sight?

5.1.2 Multisensory integration, memory and visual impairment

As discussed in Chapter 1, people with congenital blindness are in the minority within the blind population, constituting less than 5% (RNIB, 2012). The majority of AD users are therefore likely to have some sight or retain some visual memory that may be triggered by sound. Imagery has long been thought to facilitate memory (e.g. Paivio, 1971). Ogden and Barker (2001, cited in Cattaneo & Vecchi, 2011) found that LB were more likely to report visual images when asked to recall childhood memories, but drew on auditory and tactile imagery for more recent events. The interviews in Chapter 3 showed there was individual variation in the importance attached to visual memory by LB participants, with some losing interest in the visual world. As ID001 expressed it:

> "I've been totally blind for 10 years, and I was thinking of the people who were at this party yesterday lunchtime, very few of them can I actually conjure up their faces... One or two very old friends I can maybe sort of have a sense of he's got craggy features or he's tall and skinny or whatever, but that's gone really pretty well, because I don't experience people in that visual way. My reality is a reality without visual images."

Hollins (1985) suggests that vividness of visual imagery diminishes in line with the proportion of life a person has lived without sight, although this was not supported by the Study 1 participants, nor by those discussed in the article referred to in Chapter 1 by Oliver Sacks (2003).

Eardley and Pring (2006) found no difference in the number of memories generated by B&PS in response to cue words in 3 specific modalities: vision, sound and touch. This suggests that words with supposedly purely visual connotations (e.g. rainbow, clown, sunset) may trigger crossmodal or perhaps emotional associations in the blind that compensate for the lack of direct visual experience. This reinforces the question of how easy it is to isolate the modality of imageability. For example, Eardley and Pring's list of visual words included "poster" which may seem extremely visual to the sighted but has a shape and texture perceptible by touch. A memory linked to the occasion on which the poster was purchased, such as attending a gig or a concert, may be triggered by knowledge of what the poster displays, even if the visual content of the poster is not directly accessible. One blind man values holiday photographs because, when told what they show, he remembers the sounds and smells and experience of being in that particular place (G. Griffiths, personal communication, September 2012). Lenci et al. (2013) note the similarity of language between sighted and congenitally blind people, even for supposedly "visual" words such as colour terms, and conclude that useful knowledge about the visual world can be gained through other modalities.

Pietrini, Ptito and Kupers (2009) suggest that semantically-triggered imagery is supramodal irrespective of vision characteristics. This is in line with an argument put forward thirty years ago by Zimler and Keenan (1983) who found that in a paired-word association task, B, like S, remembered more pairs of words high in visual imagery than in auditory imagery; B recalled as many words grouped by colour as their S peers; and were equally compromised on a recall task when imagining objects that were “concealed” behind other objects. Zimler and Keenan concluded that either semantic representations in the blind were as effective as visual imagery for the sighted, or that the sighted were not, in fact, using visual imagery. An alternative explanation is that the sighted were not, and do not, use *solely* visual imagery. However, given what has been discussed about occipital recruitment for speech processing in B but not S, the difference (or lack of it) may be also explained through the linguistic compensation hypothesis proposed in Chapter 1. For example in neuroimaging studies by Burton and colleagues (2002a; 2002b) blind participants were asked to generate verbs from nouns or listen to nouns. Patterns of activation in the visual cortex were found to be similar to activation patterns in sighted people when presented with visual stimuli.

5.1.3 Imagery cued by verbal versus non-verbal sound

Compared with verbal cues, evidence from research in radio advertising suggests that, for sighted people, imagery is more vivid when triggered by non-verbal auditory stimuli i.e. sound effects (e.g. Miller & Marks, 1992; Bolls, 2002). Rodero (2012) puts this down to differences in the way semantic and perceptual stimuli are processed, such that semantically processed information is experienced as less intense. If this is also the case for B&PS, then Lopez and

Pauletto (2009) would be correct in their assumption that when verbal description covers the soundtrack, it does so at the expense of the user's imagination. The current study compared the clarity of auditory imagery in B&PS and S, together with the breadth and depth of that imagery when prompted by words or non-verbal sounds (sound effects). It aimed to test the following hypotheses:

H1: Due to auditory compensation, B&PS participants will imagine sounds with greater clarity than sighted participants

H2: Irrespective of sight characteristics, word cues will elicit less vivid imagery than non-verbal auditory stimuli

H3: B&PS participants will report unimodal rather than multisensory imagery

5.2 Method

5.2.1 Participants

139 volunteers, who received no payment for their participation, were recruited through organisations working with blind people, including the MDS and VocalEyes, and through the social networking site Facebook and personal contacts. Acuity of sight was assessed by questions designed for the Network 1000 study (Douglas et al., 2006). The sample was divided into 3 groups: Blind people (B) who had little or no light perception, including 5 CB (N = 18, male = 13, age 27 – 73 years, m = 50.76 yrs, SD = 13.74). Partially sighted people (PS) with some useable vision covering a wide range of sight levels. Some could just make out the shapes of furniture in a room; others could see well enough to recognise a friend at arms length, but not further away (N = 21, male = 13, age

21 – 73 years, m = 50.52 yrs, SD = 12.16). The third group was Sighted (S) (N = 100, male = 45, age 18 – 76 years, m = 46.80 yrs, SD = 13.74) with normal, or corrected normal vision.

5.2.2 Materials

5.2.2.1 Verbal stimuli

Another reason for the small number of studies devoted to auditory imagery may be the paucity of validated self-report measures. Willander and Baraldi (2010) claim these are limited to three: the original Betts Questionnaire upon Mental Imagery (1909), the shortened Betts (Sheehan, 1967) and Gissurarson's Auditory Imagery Scale (AIS, 1992). Concerned that existing measures conflate vividness and clarity, Willander and Baraldi developed their own scale focusing on the latter. While they do not specifically define auditory clarity, they refer to the distinction drawn in visual imagery by McKelvie (1995, cited in Marks, 1999) dividing vividness into clarity (brightness and sharpness) and liveliness (measured on a continuum from lively to flat). By concentrating on clarity, Willander and Baraldi aimed to limit potential ambiguity. Ambiguity also arises, they argue, from the uneven gradation in the labelling of points on the Likert scales used by existing measures. The Betts, for example, jumps from "not clear or vivid but recognisable" (level 4) to "vague and dim" (level 5). Gissurarson's AIS has an equally sudden shift from levels 1 and 2, differentiating between different degrees of "clear", to level 3 which is labelled "vague". The CAIS, by contrast, asks participants to rate clarity of imagery using a 5-point scale only anchored at each end: 1 = not at all, 5 = very clear.

Willander and Baraldi also criticize the items on earlier measures, as there is no explanation as to how the sounds were selected. The CAIS comprises written labels for 16 familiar sounds (Table 5.1) chosen to “reflect a range of common auditory experiences” (Willander & Baraldi, 2010, p.786). The CAIS has a Cronbach’s alpha of .88, which is above the .80 reliability threshold recommended by Field (2009). Having trialled the CAIS with 212 students at Stockholm University, Willander and Baraldi have called for further validation and development.

“How clearly can you imagine the sound of…?”			
Set A		Set B	
1	…a clock ticking	9	…a group of people chatting
2	…a car ignition	10	…dinner plates colliding
3	…a mobile phone ringing	11	…the rustle of a bag
4	…a dog barking	12	…a can of soft drink being opened
5	…water splashing from people bathing	13	…birds singing
6	…a person laughing	14	…the rustle of leaves underfoot
7	…the operation of a washing machine	15	…wind blowing in the trees
8	…someone sneezing	16	…paper being torn apart

Table 5.1 The CAIS (English version taken from Willander & Baraldi, 2010)

5.2.2.2 Non-verbal stimuli: Sound effects

For the purposes of this study, the 16 word cues of the CAIS were also recreated as sound effects (SFX). 14 of these were recorded by the researcher, in stereo, using an Edirol R-09HR, a 24 bit, 96 kHz wave/MP3 recorder. The remaining two sounds, Sound 5 (water splashing from people bathing) and Sound 12 (a can of soft drink being opened), came from a free online SFX site (www.sound-effects-library.com). The edited sounds ranged in duration from 14s to 17s.

5.2.3 Procedure

Participants took part online. The survey, hosted by Qualtrics (http://www.qualtrics.com/), was accessed via a unique link for each participant. It could not be completed more than once, nor could the link be forwarded to others. Those with a visual impairment accessed the written elements of the survey via a screen magnifier or by using screen-reading software, such as JAWS. The 16 cues were presented in 2 conditions, to which participants within each sight group were allocated at random. In condition 1, cues 1 – 8 (Set A) were presented as words and cues 9 – 16 (Set B) as SFX. In condition 2 this was reversed. After each verbal description, participants were asked to rate how clearly they could imagine the sound, and then how strongly it brought to mind associated imagery in each of the 7 modalities: sight, sound, touch, organic (internal bodily sensation), smell, taste and movement. For example:

"When imagining the sound of a clock ticking, how strongly does it bring to mind any or all of the following:

i. A mental picture
ii. Other associated sounds
iii. How it might feel on your skin
iv. An internal bodily sensation
v. A particular smell or smells
vi. A particular taste or tastes
vii. A sensation of moving your body?"

Cues were rated on a scale of 1 – 5 (1 = "not at all strongly"; 5 = "very strongly") for each modality. Participants could take as long as they liked to complete their response for one cue before moving on to the next. After the 8 word cues, participants responded to the 8 SFX cues. Having listened to a sound, participants were asked to identify it by a brief description and then report strength of associated imagery, using the same questions listed above. Each sound was initially triggered automatically but could be listened to again, as many times as required, before a participant moved on. Participants were also asked to rate familiarity for all cues: words and SFX (1 = "not at all familiar"; 5 = "very familiar").

5.3 Results

5.3.1 Clarity of auditory imagery: Word cues

CAIS scores (means and medians) are shown in Table 5.2. Almost half the sample (70 out of 139) reported the maximum CAIS mean of 5. The proportion was similar in each sight group, S: 50/100; PS: 10/21; B: 10/18 (including 3/5 CB participants). A Chi-square test showed no significant difference in scores between the stimuli in Set A (m = 4.53) and those in Set B (m = 4.47), p = .676. Scores were combined to provide a CAIS group mean for all 16 word cues. As the results were not normally distributed, non-parametric independent samples tests were used. These showed no significant differences between sight groups in clarity of auditory imagery: B/PS: $U = 173.5$, z = -.470, p = .666; B/S: $U = 872$, z = -.225, p = .822; PS/S: $U = 965.5$, z = -.618, p = .537.

There was a significant correlation between how clearly participants could imagine a sound and its familiarity ($r = .775$, $p < .001$). However, there were no significant differences in familiarity between sight groups: $\chi^2 = .753$, $p = .686$.

	Set A mean (SD)	**Set B mean (SD)**	**Total CAIS score mean (SD)**	**Total CAIS score median**
Blind	4.81 (0.47)	4.17 (1.34)	4.53 (0.98)	5.00
Partial Sight	4.71 (0.39)	4.52 (0.58)	4.61 (0.50)	4.88
Full Sight	4.68 (0.46)	4.72 (0.47)	4.70 (0.47)	4.93

Table 5.2 CAIS means (SD) and medians by sight group

5.3.2 Strength of imagery in each modality

Non-parametric, Mann-Whitney tests compared groups for strength of imagery in each of the 7 modalities for both word and SFX cues. This revealed significant between-groups differences only in the visual modality. Compared with both other groups, the B group had significantly lower scores whether imagery was triggered by words or SFX. Word cues: B (mdn = 3.69) compared with PS (mdn = 4.63) $U = 62.5$, $z = -3.005$, $p = .002$; B compared with S (mdn = 4.68) $U = 405$, $z = -3.116$, $p = .002$. SFX: B (mdn = 3.44) compared with PS (mdn = 4.63) $U = 66.5$, $z = -2.841$, $p = .004$; B compared with S (mdn = 4.63) $U = 319.5$, $z = -3.826$, $p < .001$. There was no difference between the PS and S groups for either cue type: Words $p = .373$; SFX $p = .564$. Figs. 5.1 – 5.3 show mean strength of imagery in the visual, auditory and body-based modalities in the words and SFX conditions, according to sight group.

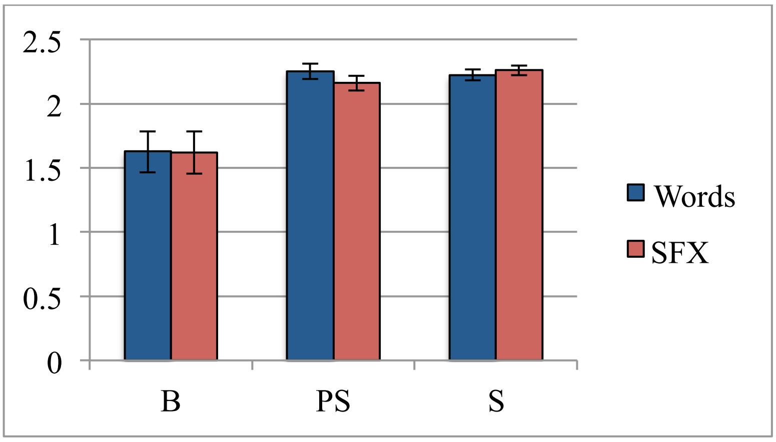


Fig. 5.1 Mean strength (SEM) of visual imagery for words and SFX

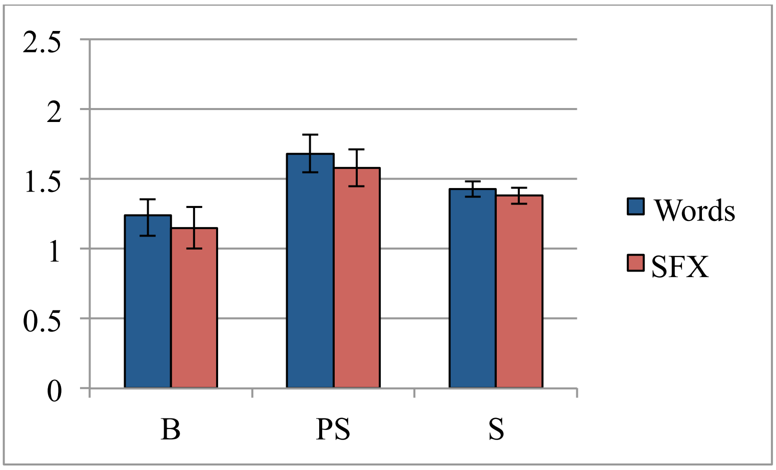


Fig. 5.2 Mean strength (SEM) of auditory imagery for words and SFX

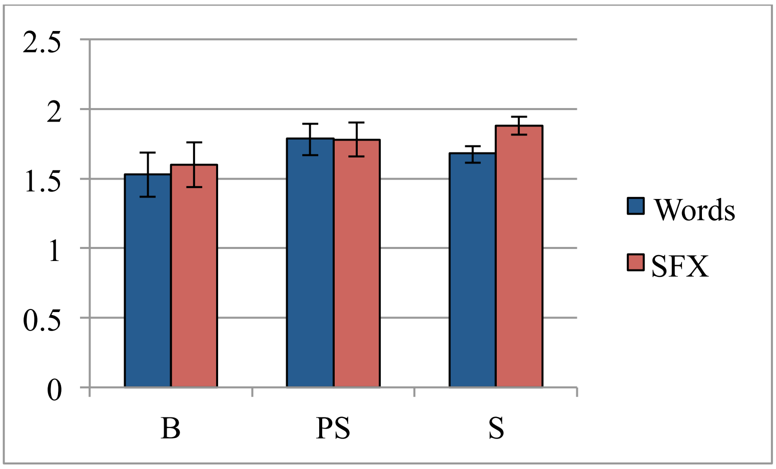


Fig. 5.3 Mean strength (SEM) of body-based imagery for words and SFX

Body-based imagery was calculated as the mean score for imagery in the modalities of touch, organic, taste and movement. Paired sample t-tests found that only the S group showed a significant difference between conditions, and only for body-based imagery: $t(96) = -6.07$, $p < .001$. This was significantly stronger when triggered by SFX.

5.3.3 Accuracy in identifying SFX

The accuracy of identification of SFX cues by participants was scored as follows: 3 = highly accurate; 2 = partially accurate; 1 = inaccurate e.g. in response to hearing Sound 7 ("the operation of a washing machine") the answer "a washing machine" scored 3; "kitchen appliance" scored 2; "train" scored 1. Scores were summed across the 16 cues. 6 participants with particularly low accuracy scores ($z > -1.96$) were excluded from the remaining analysis (B = 2; PS = 2; S = 2). Mann-Whitney tests showed no between-groups differences: B (mdn = 22) compared with PS (mdn = 22) $U = 139$, $z = -.44$, $p = .660$; B compared with S (mdn = 22) $U = 748.5$, $z = -.231$, $p = .817$.

5.3.4 Contributing modalities

Imagery scores from all contributing modalities were summed for each cue. The lowest possible score in each of the 7 modalities was 1 (not at all strongly) and the maximum was 5 (very strongly) giving a range of 7 - 35 (not at all strongly in any modality – very strongly in all modalities). This resulted in the following medians for each sight group: B words 15.75, SFX 15.5; PS words 18.31, SFX 17.44; S words 16.5, SFX 17.75. A Friedmann test showed that sighted participants reported significantly greater strength of imagery prompted by SFX than by word cues, χ^2 (97) = 6.58, p = .010. There was no significant difference between cue-type for either group with a visual impairment.

5.3.5 Breadth of Imagery

The number of modalities contributing to each image was compared. Scores ranged from 0 – 7 whereby 0 = imagery not at all strong in any modality, and 7 = at least some strength of imagery in all modalities. This resulted in the following medians for each sight group: B words mdn = 3.94, SFX mdn = 3.38; PS words mdn = 4.12, SFX mdn = 3.75; S words mdn = 3.56, SFX mdn = 3.94

As the results for the S group were normally distributed and met the assumption of sphericity, a mixed measures ANOVA was used. This showed that imagery for S was triggered in significantly more modalities by SFX (m = 4) than by word cues (m = 3.76): F (1,89) = 7.38, p = .008. However, for participants with a visual impairment there was no significant difference between the two conditions.

5.3.6 The dominant modality

The dominant modality (i.e. the modality given the highest strength rating) was determined for each of the 16 items in each condition (Figs. 5.4 and 5.5). S and PS reported strongest imagery in the visual modality for all word cues (16/16). For B (CB/LB) the visual modality was ranked top for 9/16 but for the remaining 7 word cues dominant modalities were as follows: auditory (2/16) for "the sound of a mobile phone" and "birdsong"; tactile (2/16) for "water splashing from someone swimming" and "the wind blowing in the trees"; olfactory (2/16) for "the operation of a washing machine" and "a soft drinks can being opened"; and kinaesthetic (1/16) for "the rustle of leaves underfoot".

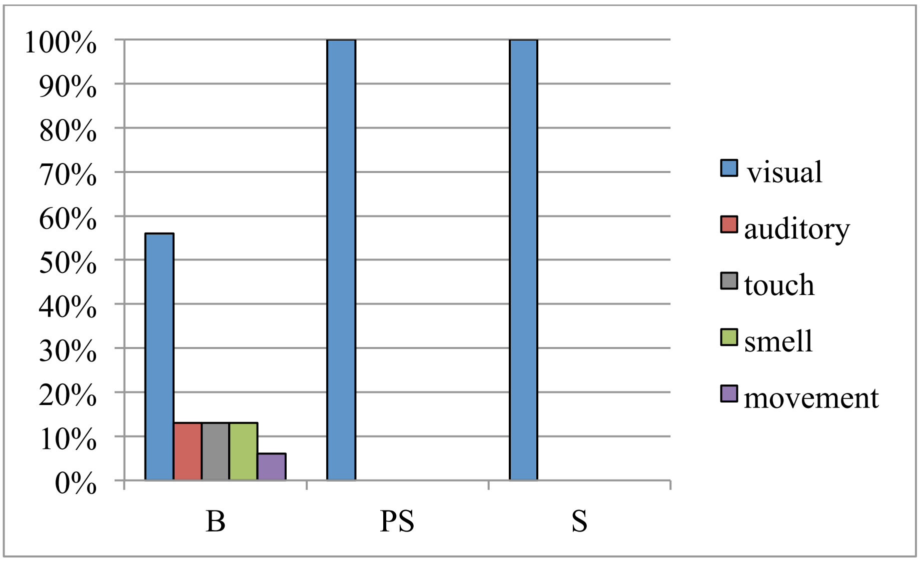


Fig. 5.4 Dominant modality for word cues by sight group

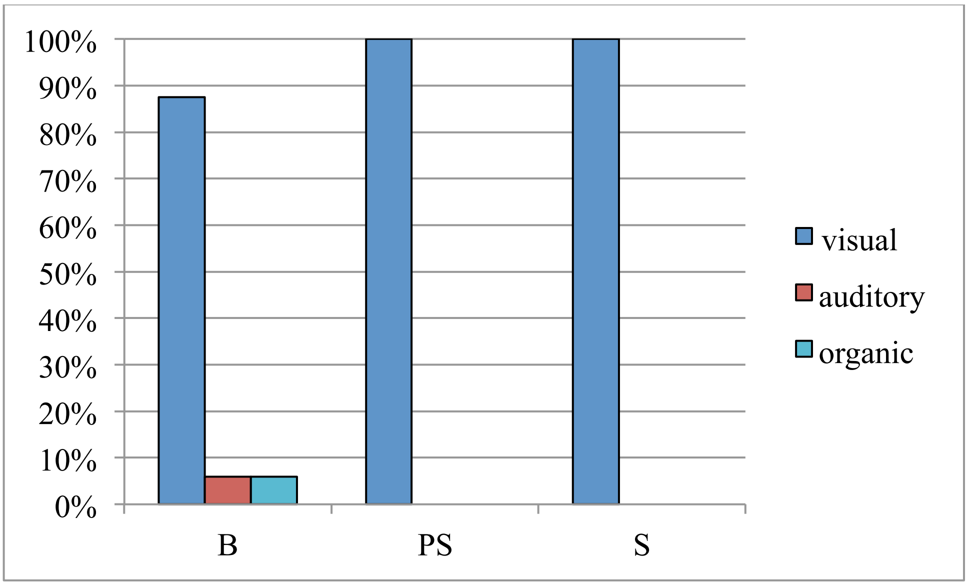


Fig. 5.5 Dominant modality for SFX cues by sight group

For S and PS, visual imagery was again dominant for all 16 cues prompted by SFX. For B, visual imagery was dominant for 14/16. For the remaining 2 cues the dominant modalities were auditory ("birdsong") and organic ("tearing paper").

5.4 Caveats

The online procedure meant there was no control over the length of time participants took to respond to each cue, nor of the quality of sound reproduction. The advantage was that B&PS participants were able to complete questions independently, without the need for a sighted intermediary. It also made it easier for this hard-to-reach population to take part. All imagery questionnaires rely on self-report, resulting in explicit, rather than implicit responses. While all participants were specifically prompted to consider associated imagery, they did not consistently report imagery across all modalities. This suggests the prompting did not, of itself, trigger their responses.

5.5 Discussion

The aim of this study was to test whether people with impaired vision have a compensatory advantage in the auditory modality, in particular for auditory imagery. It compared S and B&PS people in clarity of auditory imagery prompted by word cues (the CAIS). Furthermore, it compared the nature of that imagery: whether it was unimodal or multisensory, and whether mental representations were richer (i.e. stronger and drawing on a larger number of modalities) if stimulated by non-verbal sounds rather than words. There was no support for H1: that B&PS imagine sounds with greater clarity than sighted participants. There was partial support for H2, in that word cues elicited less vivid imagery than non-verbal auditory stimuli, but only for S and not for B&PS participants. There was no support for H3, that B&PS participants would report unimodal rather than multisensory imagery. All participants reported multisensory imagery although the relative weighting of modalities varied according to sight characteristics and condition. All these findings are discussed in more detail below.

5.5.1 Clarity of auditory imagery: The CAIS

This study found no evidence that B&PS people can imagine sounds more clearly than sighted people. However, auditory imagery is notoriously difficult to assess, and the measure used here, the CAIS, proved problematic. At least half the sample reported the ceiling score of 5, and this distribution of scores was consistent across sight groups. Marcell et al. (2007) found similarly high scores using Gissurarson's Auditory Imagery Scale and suggest this is a feature of auditory imagery i.e. people may not be able to identify subtle gradations in auditory imagery generation, but rather regard it as a binary task (can/cannot imagine).

The CAIS was developed in Sweden and it is interesting to wonder whether there is a cultural difference between this UK-based sample and the original Swedish sample that showed no such ceiling effect. Translation of the word cues is one issue and, with no definition of auditory clarity, Willander & Baraldi further rely on a shared semantic understanding of the term by their subjects. It would be interesting to substitute "clearly" with "easily", "strongly" or "vividly" to see how it affected responses.

Marcell et al. (2007) claim that recall of auditorily-presented stimuli is affected by familiarity of the sound. The current study suggests this is also true when imagining sounds, given that familiarity was so highly correlated with scores on the CAIS. As the sounds of the CAIS were deliberately chosen for their familiarity, this may be a fundamental flaw in its design. On the other hand, it would be impossible to imagine clearly a sound with which you were unfamiliar.

5.5.2 The CAIS and specificity

Interestingly, the process of creating the SFX revealed further difficulties with the CAIS. Several sounds were open to interpretation. Sound 11 ("the rustle of a bag") could differ according to the material from which the bag is made: a plastic bag would sound very different from a paper bag. It is also hard to know that the noise is made specifically by a bag, rather than, say a sheet of plastic or paper. Sound 7 ("the operation of a washing machine") varies according to the stage of the wash cycle. The word cues for Sound 6 ("a person laughing") and Sound 8 ("someone sneezing") do not specify gender, although this was identifiable when the actual sound was heard (a man in this study). Similarly from the sound of laughter, you also gain a better idea of the age of the man and how amused he is. A word cue leaves these aspects vague. Sound 12 ("a can of *soft* drink being opened" (my italics) presents the opposite problem. The sound of a can being opened is identical, whether the drink it contains is soft or alcoholic. In all, sounds 1 – 4 and 6 – 10 were more specific as SFX; Sounds 5, and 12 – 16 were less specific. Sound 11 was arguably both more specific (you could hear the bag was made of plastic) and less specific (it was hard to tell it was a bag).

While a few items on the CAIS are simple, repeated tones (e.g. a clock ticking) most are better described as "sound events" defined by Marcell et al. (2007, p. 561) as "lengthier sequences of closely grouped and temporally related sounds".

In the case of Sound 5 (“water splashing from people bathing”) the noise of water and of people must be combined, in order to cover the full dimension of what is described. The people presumably have to be making vocal sounds, to clarify that the splashing is not the result of an animal or mechanical source. In the English translation it is also unclear whether “bathing” refers to swimming or taking a bath. This item retains elements of ambiguity as a sound effect, with respondents identifying the splashing of water, but unsure as to the cause.

The items of the CAIS were not intended to be recreated as sounds, but simply as verbal prompts to generate imagery. Perhaps it is not essential that all participants imagine the same sound. If we accept that our perception of the world will vary according to experience and individual characteristics, we must accept that our imagery will also differ. There is no guarantee when completing the Betts QMI, for example, that participants asked to imagine the “smell of a badly ventilated room” conjure up the same musty odour. However, specificity is important because it allows a mental image to be more easily formed. Wyer et al. (2008) suggest that a phrase such as “a boy kicks a ball over the garden wall” prompts a stronger image than “a boy kicks a ball”, which in turn prompts a stronger image than “a boy owns a ball”. In the current study the number of items that were more specific as SFX (9/16) was higher than the number that were less specific as SFX (5/16). This may have affected within-subjects differences between the two conditions, although it does not explain the difference in performance between B&PS and S.

Arguably the CAIS itself reflects unconscious visual influence on the part of the researchers who developed it. Certainly the SFX highlight the problem of trying to design imagery measures focused on one modality in isolation. This is clearly illustrated by Sound 12 ("opening a can of soft drink"). 20% of participants identified it as opening a can of beer. Their imagery must have been drawing on modalities other than sound, as only visual, olfactory, or gustatory information would indicate what type of drink the can contains.

5.5.3 Strength and breadth of imagery: Words versus SFX

5.5.3.1 Sighted participants

Although the cue words in this study were specifically designed to prompt auditory imagery, S reported that their mental representations were predominantly visual. Williams, Healy & Ellis (1999) showed that S retrieve more autobiographical memories when prompted by word cues high in visual imageability compared to other modalities. However, comparing S and B, Eardley and Pring (2006) found no difference in the number of memories retrieved whether word cues were high in visual or auditory imageability. The findings of this study reinforce their results and provide a tentative explanation: that memories high in auditory imageability will, for those with visual experience, also trigger a visual image. Conway (2001) suggests that information is only encoded, and added to the event-related knowledge base, if it is salient. It is perhaps not surprising, then, if for those with access to visual memory, vision predominates in imagery whatever the supposed modality of the cue. It seems that imagery is inherently multisensory but for sighted people vision usually predominates, as is the case with perception itself.

In line with previous studies (e.g. Miller & Marks, 1992; Bolls, 2002; Rodero, 2012) sighted people reported imagery as more vivid i.e. significantly stronger and in a greater number of modalities when prompted by SFX than by a verbal description. Those who actually listened to the rustling of leaves underfoot, for example, reported not only visualising the scene, but also imagining associated sounds, smells and movement. Barsalou (2010), presenting the case for Grounded Theory, argues that internal representations are simulated in the brain's modal systems; that mental images are modal images. That modal systems become active in memory and thought has been corroborated by numerous neuroimaging studies (e.g. Thompson-Schill, 2003; Pulvermüller, 2005; Martin, 2007; Bhattacharya, 2009). Rodero (2012) suggests that, as memory is perceptually encoded, it will be activated more vividly by a sound cue then by a verbal stimulus, which may bypass perceptually-encoded memory to activate less vivid, semantic recognition.

5.5.3.2 Blind and partially sighted participants

Like S, PS consistently reported that a mental picture was what word cues and SFX brought most strongly to mind. This is consistent with findings (e.g. Cattaneo et al., 2010) that show mental imagery, in particular the ability to manipulate visual imagery, is similar in sighted people and those who have had at least some visual experience. However, PS also reported significantly less difference between word cues and sound cues for strength of mental representation than their fully-sighted peers. This supports the suggestion that difference in imagery is based on encoding. Goddard et al. (2005) argue that, for people with a visual impairment, memory is more likely to be verbally mediated.

People with a visual impairment may be habituated to the congruence of sound and a verbal description, rather than to the congruence of sound and (unreliable) vision.

In this study, B participants (which included CB and LB) reported imagery that was significantly less rich in visual imagery compared to PS and S. It is likely that this would have been even more pronounced had the sample only included CB participants. B participants also reported a variety of modalities as dominant, depending on the particular stimulus. For example, smell was triggered more strongly than touch when imagining the sound of opening a can of soft drink. Clearly it is important to identify the contents of a can, which cannot be done by sound alone. Smell compensates for the absence of sight, saving the blind drinker from a potentially unpleasant, or at least unexpected, gustatory experience. By contrast, touch was rated more strongly than smell when imagining the sound of someone swimming. Given the comments on embodied imagery reported in Chapter 3, it is likely that some B participants, not being in a position to "observe" others in their mind's eye, instead imagined swimming themselves. This may explain why "how it feels on your skin" was what the sound brought most strongly to mind, rather than associated sensations such as the smell of chlorine or the taste of brine. After completing the experiment, one participant (ID014) explicitly stated: "I think all [my mental images] were how I would picture me…or a memory of me in that situation".

It is interesting that, for B, the visual modality was more dominant for SFX cues (14/16) than word cues (9/16). As in Study 2, it is possible that visual imagery in those with visual experience was prompted by auditory cues, reflecting bimodal associations that persist after sight loss. However it is also possible that the term “mental picture” did not in fact denote visual but spatial information for some participants. In this case perceptual sounds (but not words) may have helped to build a stronger “picture”, regardless of whether that “picture” contained any information about visual appearance. Spatial mental models will be discussed in the next Chapter and explored more fully in Chapter 7.

There was no advantage for B&PS participants when it came to identifying non-verbal sounds. In addition, the verbal description proved at least as rich, in terms of strength of imagery and number of modalities evoked, as listening to the sound itself. There was no support, then, for the initial element of H2: *Irrespective of sight characteristics*, word cues will elicit less vivid imagery than non-verbal auditory stimuli. This was only the case for sighted people.

There was also no support for H3: that B&PS participants will report unimodal rather than multisensory imagery. Indeed, the findings counter suggestions that vision is a prerequisite for multisensory integration and support evidence of crossmodal associations in blind individuals. Théoret, Merabet and Pascual-Leone (2004) have suggested that the visual cortex is, in fact, multisensory but its crossmodal connections are “masked” by the dominance of vision. This will be returned to in later Chapters. The findings also strengthen the possibility that language can facilitate crossmodal integration in the absence of sight. In contrast to the implicit crossmodal integration that vision allows, B&PS have a greater reliance on language to create a coherent understanding of the world.

5.6 Summary of Study 3

Study 3 suggests that heightened auditory perception in people with a visual impairment neither extends to the ability to identify sounds out of context nor strengthens auditory imagery, although this latter may be the result of a ceiling effect in the CAIS. Study 2 showed that auditory/haptic associations are indirectly influenced by vision. Study 3 suggests that vision also shapes imagery arising from real or imagined auditory cues. However, it also highlights the problem of trying to consider imagery in any modality in isolation. All participants reported multisensory imagery, although for sighted people it was less multisensory when prompted by a verbal description. For people with impaired vision verbal and non-verbal sounds equally sparked the imagination.

Before moving on, it is important to draw attention to a couple of significant caveats. By asking participants to identify the SFX before reporting associated imagery, it is possible that they were responding to their own description, as much as to the SFX themselves. It may be argued, then, that in both conditions, the stimuli were semantic. Although this would not explain the significant increase in strength/breadth in sighted people's imagery in the SFX condition, this aspect is addressed in the next study.

As far as AD is concerned, the findings suggest that obscuring information from the non-verbal soundtrack with a verbal description will not necessarily diminish the richness of mental imagery experienced by AD users. However, given the comments on flow reported in Chapter 3, it is hard to extrapolate from imagery responses that are allowed to develop over time to those in a dynamic environment, such as a play or a film, where sound stimuli are constantly being updated. The next study addresses this by testing responses to scenes from an audio drama. Responses are measured with the ITC-Sense of Presence Inventory (ITC-SOPI, Lessiter et al., 2001) to compare the effects of words (dialogue) and non-verbal sounds (SFX) on presence.

Chapter 6 Study 4: Audio Drama: Sound, Words and Imaginal Presence

6.1 Introduction

So far it is apparent that, although visual experience influences bimodal associations, it is not essential for an individual to enjoy multimodal imagery. That imagery may contain a visual component depending on an individual's sight characteristics. Regardless of those characteristics, however, for B&PS imagery can be prompted as effectively by words as by non-verbal sounds. This is potentially important for AD, which has been criticised for obscuring the non-verbal soundtrack of its source material. This consists of music and sound effects (SFX), either natural or synthesized, which will be associated with particular actions, objects or environments.

A purely auditory medium ranks low on the scale of immersive environments. Yet there is a popular belief that, for sighted individuals, "pictures" are better on radio and that mental images evoked by sound create a "theatre of the mind" (Crook, 1999, p. 61). As discussed in Chapter 2, media form is just one determinant of presence. Media content is also important. Murray (1997, pp. 98-99) argues that "a stirring narrative in any medium can be experienced as a virtual reality because our brains are programmed to tune into stories with an intensity that can obliterate the world around us".

In a previous study (Fryer & Freeman, 2012a; 2012b) based around a clip from David Lean's black and white classic film *Brief Encounter*, B&PS watching with

AD reported levels of presence that were similar to, or higher than those reported by S watching without AD (this will be discussed more fully in Chapter 8). S reported lower levels of presence when the clip was audio described. This was explained by conflicting processing demands between the verbal commentary and the visual stimulus. For S the AD appeared to interfere with visually-processed information, whereas B&PS could better shape their own mental representation with no conflict from the images on screen. However, if the results of Study 3 also apply to dynamic scenes, it is possible that the verbal commentary was less evocative for sighted people, compared with SFX and music, while for B&PS both were equally effective.

A study by Rodero (2012) explored sighted people's responses to scenes from an audio drama, presented in two conditions: with dialogue alone and with dialogue plus SFX and sound shots (i.e. a difference in intensity of sound, induced by the position of the microphone, that gives a spatial effect equivalent to foreground or background in a visual image). Sighted listeners reported more images and greater attention for the scenes with SFX. If Rodero's findings extend to B&PS, then it may be better for the describer to be silent at times when sound effects can speak for themselves, allowing the listener to create a more vivid mental representation.

A pilot study compared levels of presence in B&PS and S when listening either to the dialogue of an audio drama, or to a combination of dialogue and SFX. It suggested that the addition of SFX increased ratings on some dimensions of presence for S, but not for B&PS. Using a longer stimulus (3 scenes lasting almost 6 minutes compared to a single scene lasting just over 1 minute) the study reported here was designed to see if the results of the pilot would be replicated. The aim was to test two hypotheses:

> H1: S participants listening to scenes with SFX would report higher levels of presence compared to listening without SFX.

> H2: B&PS participants would report similar levels of presence, whether SFX were present or absent.

6.2 Method

6.2.1 Participants

Participants volunteered to take part for no payment. All were known to the researcher or had been contacted via organisations that support people with sight loss e.g. RNIB and the MDS. Participants were asked to identify themselves as having no, some or full useable vision. Those with a visual impairment reported the age at which they were registered as blind or partially sighted. Visual acuity was designated more specifically, using questions devised for the Network 1000 study (Douglas et al., 2006). 3 were excluded for incomplete data, leaving 21 B&PS participants (male = 16, age 29 – 70 yrs, m = 48.33 yrs, SD 11.24) comprising CB/EB (3); LB (11) and PS (7). Of the 51 sighted volunteers, 21 were selected to match the B&PS group as closely as possible in age and gender (male = 16, age 24 – 74 yrs, m = 48.43 yrs, SD 10.70).

6.2.2 Stimulus

Participants listened to a sequence of three scenes from an audio drama. The sequence, lasting 5m 56s in total, was taken from *An Everyday Story of Afghan Folk*, first broadcast on BBC Radio 4 in August 2012. It was written and directed by Liz Rigbey with sound design by David Chilton. Scenes 1 and 3 were set in a small house in a village on the Pak-Afghan border; Scene 2 took place in the fields nearby. In the version with SFX, the general location was established at the beginning by a 15s burst of music mixed with general animal noises – mooing, the sound of goats and cattle-bells. A creaking door indicated the move inside, with the sounds of cutlery and a crackling fire, goat noises and the distant barking of a dog. The transition to Scene 2 was marked by a music “sting” (dur: 7s) and the cawing of crows. The scene was underscored by birdsong with occasional SFX of goats, buzzing insects, barking and distant gunshots. Another 7s sting combining music with the sound of footsteps took us back inside the house for the final scene, with similar SFX as for Scene 1.

6.2.3 ITC-SOPI

Presence was measured using the ITC-SOPI (Lessiter et al., 2001). As discussed in Chapter 2, participants rate 44 statements on a 5-point Likert scale (1 = *strongly disagree*; 5 = *strongly agree*) with a mean of 2.5 deemed the threshold for experiencing presence. The ITC-SOPI was designed for immersive AV media. For the purposes of this audio-only study, and given the nature of the participants, the measure was adapted. One question (“I felt I had eye strain”) was removed. Another (“I felt I wasn’t just watching something”) was rephrased as “I felt I wasn’t just listening to something”. For all questions that used the phrase “in the displayed environment”, the words “displayed environment” were replaced by the words “clip” or “scene”.

The ITC-SOPI assesses 4 dimensions of presence: sense of physical space or Spatial presence (e.g. “I felt characters or objects could almost touch me”); Engagement (e.g. “I felt involved in the scene”); Ecological validity (e.g. “the scenes seemed natural”) and Negative effects (e.g. “I felt I had a headache”). This allowed for any differential impact of SFX on the various constituent elements of presence to be assessed separately, rather than providing a blanket measure of presence as a whole.

6.2.4 Procedure

Participants from each sight group (S/B&PS) were allocated to 2 groups at random: Group 1 experienced scene 1 with dialogue only and scenes 2 and 3 with dialogue *and* SFX. For Group 2, conditions were reversed. After hearing each stimulus, respondents completed the adapted version of the ITC-SOPI (Lessiter et al., 2001). The questions and stimuli were presented via an online survey site, Qualtrics (http://www.qualtrics.com/). Participants completed the experiment at home, using their own computer. Those with impaired vision used accessible screen-reading software if necessary.

6.3 Results

6.3.1 Note on analyses

The 4 presence subscales of the ITC-SOPI cannot be combined to make a single presence score (Lessiter et al., 2001). They were therefore analysed separately.

Although the subscales are part of the same construct, suggesting MANOVA as a suitable statistical test, Levene's Test was significant in the SFX condition for Spatial presence: $F(1, 40) = 4.59$, $p = .038$; Ecological validity: $F(1, 40) = 4.69$, $p = .037$; and Negative effects: $F(1, 40) = 8.87$, $p = .005$, violating the requirement for homogeneity of variance between sight groups (B&PS/S). Sight groups were therefore analysed separately using within-groups repeated measures.

6.3.2 Sound effects and presence

Radio is a low-immersion medium and scores on all 4 presence subscales (max = 5) were consequently low. Mean ratings for each sight group (B&PS/S) in the two conditions (with/without SFX) are shown in Table 6.1. Using mid-point of the 5-point scale (i.e. 2.5) as the threshold score required to feel at least some sense of presence, neither group attained a sense of Spatial presence, although the sighted group approached it in the SFX condition. Ratings of Negative effects were also noticeably low.

Results revealed significant differences between S and B&PS participants. For S, Spatial presence, while low, was significantly higher with SFX ($F(1, 20) = 9.73$, $p = .005$). Ratings of Engagement ($F(1, 20) = 7.78$, $p = .011$) and Ecological validity ($F(1, 20) = 7.59$ $p = .012$) were also significantly higher with SFX than listening to dialogue alone. The impact of SFX on Negative effects was not significant ($F(1, 20) = 2.20$, $p = .153$). H1 was therefore supported for the three positive dimensions of presence.

	Condition	B&PS	S
Spatial Presence	No SFX	2.26 (0.86)	2.09 (0.67)
	SFX	2.25 (1.04)	2.41 (0.61)
Engagement	No SFX	3.15 (1.02)	3.10 (0.84)
	SFX	3.23 (1.03)	3.40 (0.68)
Ecological Validity	No SFX	3.04 (0.97)	3.12 (0.90)
	SFX	2.96 (0.92)	3.49 (0.64)
Negative Effects	No SFX	1.66 (0.70)	1.50 (0.59)
	SFX	1.61 (0.82)	1.40 (0.44)

Table 6.1 Mean (SD) presence scores by sight group and condition (No SFX/SFX)

For B&PS ratings were not significantly different on any of the presence subscales between listening with and listening without SFX: Spatial presence (F (1, 20) = .002, p = .964); Engagement (F (1, 20) = .264, p = .613); Ecological validity (F (1, 20) = .134, p = .718); Negative effects (F (1, 20) = .123, p = .730). H2, then, was fully supported by the results. Fig. 6.1 shows individual changes in score between the two conditions for each sight group for each of the presence measures. Blue lines indicate an increase in score from No SFX to SFX; red lines indicate a decrease in score; black lines indicate the score stayed the same.

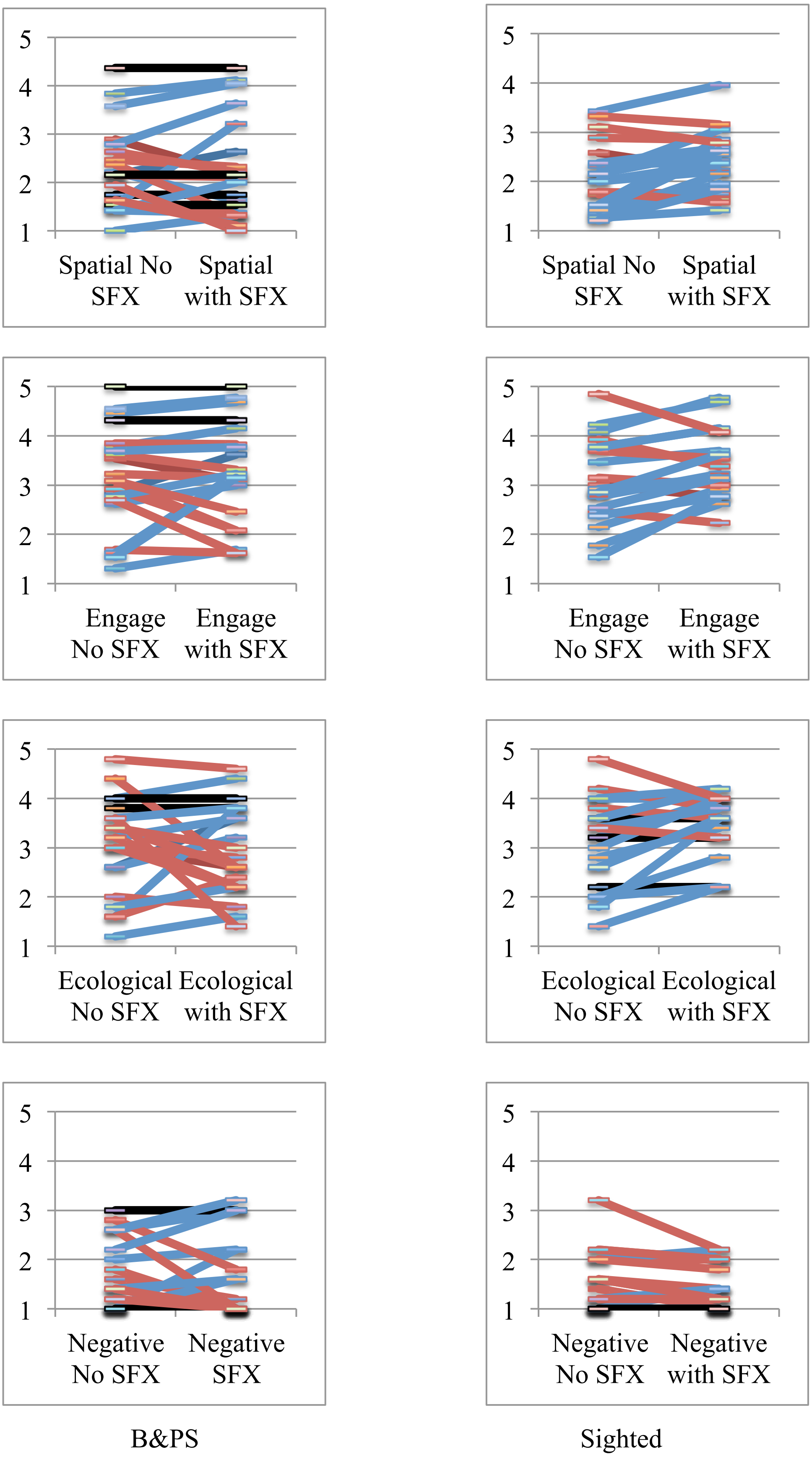


Fig. 6.1 Presence ratings by condition (No SFX/SFX) for each sight group

6.3.3 Further analyses

For the sample overall, correlations showed scores on all three positive presence subscales to be significantly linked in both conditions: Dialogue-only: Spatial presence and Engagement r = .710, p < .001; Spatial presence and Ecological validity r = .631, p < .001; Engagement and Ecological validity r = .661, p < .001; With SFX: Spatial presence and Engagement r = .622, p < .001; Spatial presence and Ecological validity r = .743, p < .001; Engagement and Ecological validity r = .622, p < .001. There were no significant associations in either condition between the positive dimensions of presence and Negative effects.

In the dialogue-only condition, for B participants (N = 14) the correlation between Spatial presence and Ecological validity was not significant: r = .464, p = .110. However there were no significant correlations between age at onset and percentage of life lived with impaired vision for any of the presence measures in either the dialogue-only or the SFX condition.

6.4 Discussion

6.4.1 Sound effects and sight

This study was small in scale, and exploratory. Participants with a broad range of sight characteristics listened in a variety of settings. However, in keeping with the results from Study 3, the findings here suggest that SFX may have a reduced impact for B&PS relative to those with sight. Rodero (2012) showed that S were likely to have a more intense imaginal experience when exposed to non-verbal compared with verbal sounds. Her findings were replicated here in terms of presence. The addition of SFX gave rise to higher ratings of Spatial presence, Engagement and Ecological validity. For B&PS people there was greater variation. Overall, however, listening to a combination of dialogue and sound effects did not make the experience significantly more realistic or immersive than listening to the dialogue alone.

The stimulus included two types of SFX: effects that illustrated the dialogue, such as audible gunshots when the protagonists mentioned target practice, and effects that supplemented the dialogue, such as the distant sound of a dog barking. The dog was not referred to by the characters but simply helped to set the scene. Sighted people, hearing the gun for example, appeared to experience the scene more vividly than those who only learned through the dialogue that target practice was taking place. In particular, they gave higher ratings of Spatial presence than other sighted people listening to dialogue alone. They were more likely to agree with such statements as “I felt I was visiting the place where the scene was happening”. They were more engaged (e.g. “I felt myself being drawn in”) and also reported stronger Ecological validity i.e. the scenes seemed more believable and more natural.

By contrast, B&PS listening with SFX reported only a small and non-significant increase in ratings of Engagement compared with those listening to the dialogue alone, and were no more likely to agree with statements indicating Spatial presence or Ecological validity, such as “I had a sense of being in the scene”, “I felt I was in the same space as the objects and characters” or “the scene seemed natural”.

6.4.2 Sound effects and dimensions of presence

6.4.2.1 Engagement

The content, which Lessiter et al. (2001) define as the theme or narrative of the drama, was enough to create a sense of Engagement through the dialogue alone for both S and B&PS listeners. Engagement increased with the addition of SFX, but only significantly so for S. The larger SDs for presence ratings for B&PS in the SFX condition overall, was particularly noticeable. This may reflect the variation in retention of visual imagery of LB, as suggested by Sacks (2003) and confirmed by interviewees in Chapter 3. For some B&PS words maybe engaging enough by themselves; for others environmental sounds may make a scene more engaging, perhaps by stimulating implicit bimodal (auditory-visual) associations drawing on residual visual memory.

6.4.2.2 Spatial presence

Regardless of condition, B&PS failed to register a sense of Spatial presence. In the dialogue-only condition, most S were also unable to generate a sense of "being there" but were on the threshold of doing so once SFX were added. It is interesting to speculate why this might be the case. Wirth et al. (2007) argue that, of all the dimensions of presence, Spatial presence is closest to Minski's (1980) original concept of telepresence i.e. the sense of being in the mediated environment. Spatial presence itself, they claim, consists of two dimensions: feeling physically located within the mediated environment (self-location) and perceiving the possibilities for action within that environment rather than in the real world. They describe a two-step "spatial situation model" (SSM) whereby the user first must be able to identify the mediated space, and then feel that they themselves are located in that space. Of particular importance is attention, which, they argue, must bound by the medium. Media form is the critical factor here, with the greater the breadth and depth of stimuli the better. As suggested in the introduction, radio is a low-immersive medium. For S, then, additional SFX may help by capturing the listener's involuntary attention, diverting them from ambient sound in their real environment that might otherwise lead to a BiP. In this way, enhanced media content compensates for low-immersion media form.

Why should this be the case for S but not most B&PS participants? King (in Spence & Driver, 2004) argues for the primacy of vision in spatial processing of an object or event that is experienced multisensorily. In particular he suggests that multisensory integration depends on spatial coincidence: visual and auditory stimuli that are close together spatially as well as temporally will be perceived as congruent by a developing infant. As previously discussed, the visual cortex in sighted people can be stimulated by a non-visual (auditory) prompt, even when individuals have previously been exposed to it as part of a bimodal (AV) stimulus for only a short time (Zangenehpour & Zatorre, 2009). Wirth et al. (2007) suggest that imagination can compensate for lack of sensory input from the mediated environment, so if the audio drama triggers visual images in the sighted user's imagination, and these are spatially "tied" to the sounds they hear, then their SSM will effectively be supported by two modalities, sound and vision. Given the findings of Chapter 5, that SFX stimulate imagery in more modalities than verbal prompts alone, the addition of SFX will increase the breadth and depth of a sighted person's SSM.

For those with a visual impairment, perceptual associations are generated between sound and other non-visual modalities, for example between sound and touch rather than sound and sight, at least for a stimulus that is within arms' reach. As discussed in Chapter 2, however, sounds of spatially distant objects are effectively unimodal and, instead of vision, will be associated with a verbal label or description. This would explain why SFX are no more effective in developing or supporting a blind person's SSM than the verbal prompt provided by the dialogue.

In the pilot study for the current experiment, B participants actually reported lower levels of Spatial presence for the scene with sound effects. This was not replicated, perhaps reflecting differences in content. In the current study the three scenes conveyed a combination of internal and external settings, and a multiplicity of sound effects for both proximal and distant objects. It is possible that B&PS literally felt distanced by exposure to more spatially distant auditory information, but less so by SFX relating to objects or actions close to hand.

6.4.2.3 Ecological validity

SFX made the scene feel significantly more realistic to S but not to B&PS. Arguably S, exposed, for example, to the distant sound of the dog barking could build the dog into their mental image, perhaps "glimpsing" it through an imagined window or open doorway of the village house. It is possible that they have seen news footage or documentary reports from Afghan villages where roaming dogs are generally part of the background scene. The sound of the dog may therefore make the scene more realistic. This is less likely to be the case for

B&PS: TV news is currently not audio described and documentaries (other than nature documentaries) rarely offer the opportunity for much background description (Ofcom, 2001).

Whether or not that is the case, in the absence of sight, audition provides information on location and direction (e.g. Klatzky et al., 2006). With less opportunity to “observe” the source of distant sounds in the mind’s eye and integrate them into their mental picture, B&PS may have been more keenly aware that they were not in the same space as the objects they could hear, even if the sounds of dogs, goats and birds were realistic. This may explain why the correlation between Ecological validity and Spatial presence was not significant for B participants. One further possibility is that, being more strongly attuned to using sound in, for example, navigation, B were more likely to identify the SFX as inauthentic.

An alternative explanation is that in everyday life, B&PS are less able to draw on exogenous visual cues such as flicker, colour, luminence or movement to guide attention towards the cause of a sound. They may therefore have to work harder to pick out sounds of particular interest. For example, in dealing with “the cocktail party problem” (Cherry, 1953) S have the advantage of lip movements and facial expression to aid them in isolating one conversation from another. B&PS do not. As discussed in the introductory chapters, some compensation may by provided for B&PS by greater capacity for divided and selective attention for auditory information. However, the SFX that supplemented the dialogue – regarded as a bonus by sighted people – may have been filtered out by

B&PS as they were not critical to the narrative, and simply added to cognitive load. Illustrative SFX are even more likely to have been regarded as redundant information. Some evidence for this comes from the comments of participants in Study 1 who, for example, found no benefit of description of sounds or actions they could deduce for themselves. As ID004 put it, when talking about a nature documentary:

> "They start to tell me about the cubs and I hear "cubs playing in the snow" – no we've just heard the bear crunching through the snow and we know it's in the Antarctic and there's bears, so we don't need to be told that. Superfluous description can be irritating actually. I have even known them to describe sounds before…"so and so bangs on the door" and you've just heard them banging on the door. I mean OK …maybe there's some need for it but I suppose I'm so used to interpreting sounds so perhaps I feel it's slightly irritating."

The audio drama presents the reverse situation, with the information provided in the dialogue seen as sufficient without the need for SFX.

6.4.2.4 Negative effects

Neither S nor B&PS reached the threshold for Negative effects in either condition. This may reflect the fact that the ITC-SOPI was designed for immersive media environments that may cause physical responses of nausea or dizziness. Given its low-immersive form, Radio drama is unlikely to induce such side-effects.

6.4.3 Audio drama, sound effects and presence

All participants, regardless of sight characteristics, reported low ratings for all four dimensions of presence. This is unsurprising given that radio is a low-immersive medium. It may also reflect non-optimal listening conditions, although this is consistent with the way audio drama is usually received; the Radio Advertising Bureau reports “90% of listeners are actually doing something else while listening to radio” (RAB, online).

For S, ratings for positive dimensions of presence were significantly higher with the addition of SFX. It is interesting that for another low-immersive medium, namely text, Gysbers et al. (2004) found that Spatial presence was higher when less detail was provided. Study 3 showed some SFX could be more specific, compared with words. In this study, the gunshots, for example, would allow you to determine more about the calibre of the weapon. However, it is unlikely that any of the participants here had personal knowledge of the location (Afghanistan) and arguably few had experience of shooting. They would therefore not be troubled by the realism (or not) of the SFX when constructing their imaginal scene.

Gysbers et al. were comparing less text with more text. Here the comparison was between less and more sound. However, the qualitative difference between language and non-verbal sound, i.e. explicit and implicit information, may have overridden the problem of specificity. Alternatively, it may simply be that SFX helped “grab” sighted listeners’ attention. Wirth et al. (2007) suggest that the trait of absorption is an important factor in whether or not an individual accepts the mediated environment as their own PERF. B&PS may find it easier to be absorbed given the lack of visual distractions from their real environment. Collignon et al. (2006) have suggested that heightened performance in auditory attention may underlie B's apparent success in out-performing S in auditory perception tasks. In the current study, without exogenous triggers for visual attention, S may have benefitted from the exogenous auditory triggers that SFX provide. If visual perception is improved by involuntary orientation to sound (McDonald et al., 2000), it is feasible the same effect may be found for imagery.

Slater and Usoh (1993) suggest that levels of presence are influenced by individual differences in encoding preferences, with those showing visual dominance over auditory or kinaesthetic dominance experiencing greater levels of presence. Yet, as discussed in Chapter 2, research shows that, in video games, sound quality seems to be more important than image quality in influencing presence (e.g. Skalski & Whitbred, 2010). Where visual information is unreliable, sighted people are used to exploring objects through other modalities. Thus, historically, a "gold" coin would be bitten to test its level of purity; a bell would be struck to find out if it will "ring true". One explanation may therefore lie in the ability of SFX to compensate for the absence of images by stimulating rich multimodal imagery. For B&PS, for whom absence of reliable visual information is part of daily reality, additional SFX provide less of an advantage.

6.5 Summary of Study 4

The results of Study 4 are consistent with Study 3. They also extend that study showing that, for sighted people, SFX not only increase the depth and breadth of imagery for *static* mental representations, but also lead to a greater sense of presence for *dynamic* scenes. That sight is a factor in determining strength of presence, even in an audio-only media form, is important for our understanding of presence. It suggests that sound enhances presence when images are poor or missing by drawing on crossmodal associations at encoding to prompt richer imaginal presence.

As SFX bring less advantage for people with a visual impairment, these results reinforce the differences between language and perceptual processing discussed in Chapter 5. Participant comments from Study 1 suggest that, even with verbal prompts, blind people imagine themselves engaged in an activity, rather than as onlookers. Arguably this is because they are used to verbal mediation in their everyday lives. By contrast, sighted people may associate words with reported incidents of which they have had no direct experience; whereas non-verbal sounds arising from the action are more likely to evoke an experience in which they have taken part themselves.

Whatever the explanation, the findings have implications for AD. Sighted people might assume that, for AD users, minimising occlusion of SFX will increase a sense of "being in the scene". However, obscuring the soundtrack for the duration of an AD utterance may not have a detrimental effect on presence, as long as the AD is giving information that is supplementary to the dialogue.

It is proposed here that SFX influence sighted people by enhancing their spatial situation model (SSM). Wirth et al. (2007) suggest that spatial knowledge and the ability to recreate that in the imagination becomes more important, the more fragmented and less “real” the media form. They also point out that individuals create such a model from perception and memory, drawing on their event-related knowledge (ERK) base. It is useful, then, to know more about how everyday experience influences the spatial mental models created by S and B&PS.

It has also been suggested here that people with a visual impairment need to filter out unnecessary sounds, and in particular distant sounds, in order to pick up information about their immediate environment. Furthermore, although in proximal space they may make implicit perceptual connections, for example between sound and touch, they will rely on explicit associations to comprehend distal space. Such explicit knowledge is likely to be held in the memory of B&PS but not S, who are able to integrate implicit cues about distal space through the modality of vision. These hypotheses are explored in Chapter 7.

Chapter 7 Study 5: The Role of Experience on Spatial Cognition: Comparing Navigation Strategies

7.1 Introduction

So far, we have been considering how language and perceptual non-verbal sounds (i.e. SFX) affect mental imagery and aspects of presence in a mediated environment depending on an individual's sight characteristics. For sighted people only, non-verbal sounds have been shown to stimulate significantly richer imagery than words, and increase levels of presence. It has been suggested that this is related to a difference between semantic and perceptual processing, which in turn reflects explicit (declarative) and implicit (non-declarative) memory (Schacter, 1992).

The transfer-appropriate processing (TAP) principle (Morris et al., 1977) suggests that successful memory performance involves reactivation during retrieval of those same areas of the brain involved in encoding, although this depends to some extent on the type of target information (Johnson et al., 1993). Comparing semantic versus perceptual stimuli, an fMRI study by Prince et al. (2005) has shown that Relational Memory (RM), which links different perceptual and semantic information from the same event, resides in the left hippocampus. This area of the brain is activated during encoding and recall of both perceptual and semantic information. However, the study also confirms the involvement of different brain regions for each type of processing, with considerable overlap of those same particular regions during recall. Although it was carried out on S, rather than B, the findings suggest a high level of explicit input at encoding may result in a high level of explicit output at recall. As outlined in the preceding chapters, neuroimaging data suggests that the occipital cortex is recruited for verbal processing in blind but not sighted people, and is also involved in perceptual processing of non-visual modalities. It is proposed that direct associations between modal information and words result from real-world experience of B&PS that is commonly mediated by language.

At the end of Chapter 6 it was proposed that adopting a mediated environment as a personal reference frame depends on building a strong spatial situation model (SSM). Gysbers et al. (2004) suggest that where spatial cues in the mediated source are sketchy, they will be supplemented by spatial attributes retrieved from memory. Discussing S immersing themselves in a book, they argue that "verbal descriptions of spatial structures do not trigger perceptual processes such as the visual identification of an edge, but rather stimulate space-related cognitions based on existing knowledge structures" (p.14). This raises two questions. How similar are the existing knowledge structures of S and B&PS in relation to spatial cognition? And, given that B&PS participants have so far shown little evidence of significant imaginal differences arising from word prompts compared with perceptual prompts, how might sight characteristics affect the way spatial information is encoded?

Ungar et al. (2004) suggest that the benefit of sight is that it provides, both pre-attentively and attentively, an overview of the space to be explored. This allows a sighted individual to pick out patterns or clusters of patterns that they can then "zoom in" on to explore more closely, or "zoom out" from to see how a particular element fits within the total picture. For sighted people this process appears to happen "automatically". Without the advantage of a clear *Gestalt* image, one of the PS participants in Study 1 (ID003) explained how he consciously uses his residual vision to achieve the same result:

> "I come into somewhere which is strange and I see a panoramic view, but I see no detail in it at all really. I might see that there's a red brick building, I might see there's a bit of scaffolding and I might see… um…I might see there's a lawn on the other side. Now if I need to look at any of those, to find a way through, then I've got to look towards the lawn, towards the building, straight ahead, where is the path? OK. So what my brain is doing all the time and what my eyes are doing all the time is...I see a panoramic view and then I have to start focusing in small chunks to rebuild the picture. To find a way out…So what my sight and my memory is doing with me is - go back to the panoramic thing and then experience to be able to plot a course. So I appear to be walking down the road fully aware of what is going on and all I'm actually doing is following a very narrow route. Right. And so I actually have a situation, where, I always remember overhearing somebody at work saying – this is 30-odd years ago – oh I don't believe he's blind, you see, he can see a lot more than he says he can see…what I'm actually able to do is use what little sight I've got very effectively and that's the difference."

By contrast, people with little or no light perception navigate a space using proximal cues. Ungar et al. (2004) point out this is like being permanently "zoomed in". Participant ID014 described her strategy in an unfamiliar space:

> "The best thing is to go round it really and actually touch where things are, because you can walk round things and not touch anything and not have a clue really. So if you really want to know and then feel independent and find your way round, the best thing is just to go round and know where the position of chairs and tables are and stairs and all that. And then I would… I'm going to say picture it in my mind, but I mean remember the positioning I suppose, that's what it is, isn't it? I suppose as I don't see it as a whole it's bitty, but it's put together in my mind somehow as me in relation to things."

The aim of this next study is to understand the impact of visual impairment on Spatial presence by exploring how B&PS and S commit spaces to memory. It focuses on how important clues to navigation, such as landmarks, are encoded through a range of multisensory cues, by quantifying the features people verbally identify when recalling a familiar route. The intention is to elicit what type of spatial information is gathered and held in an individual's event-related knowledge base.

7.1.1 Wayfinding for people with sight

Sight is presumed to be crucial to wayfinding. Gibson cites the philosopher Bishop Berkeley who, in 1709, claimed the chief role of vision was for animals "to foresee the benefit or injury which is like to ensue upon application of their own bodies to this or that body which is at a distance" (1986, p. 232). Gibson applied Berkeley's idea to his own theory of affordances such that "to see things is to see how to get about among them" (p. 223). More recently, sight has been cited as crucial for, in particular, "the ability to orient with respect to a known object or vista of a scene, which requires memory for a particular landmark or view" (Foo et al., 2005, p. 195).

Landmarks have two functions (Allen, 1999). They allow people to orient themselves in a particular direction and provide reassurance that they remain on the correct route. An alternative to a landmark-based navigation strategy is "path integration" or "dead reckoning", that uses proprioceptive information such as muscle movement, and linear and angular acceleration of the body to constantly update position and distance travelled from a known starting point. Foo et al. noted the use of such a strategy by sighted people when asked to journey across a virtual desert that had no obvious landmarks. They concluded that S can "fall back on…path integration if landmarks are absent or perceived as unreliable" (2005, p. 213).

While other senses such as sound and smell can provide information about features of the environment, non-visual modalities have been deemed to be significantly less precise (e.g. Chandler, Grantham & Leek, 1993; Waller & Greenauer, 2007). This might suggest that people with a visual impairment rely on path integration. Yet 40 years ago Foulke (1971) noted the use of landmarks by blind people. Indeed, "using landmarks and cues" is one of nine spatial concepts orientation and mobility (O&M) trainers deem necessary for B&PS students to learn (Wall Emerson & Corn, 2006). This begs the question of what constitutes a landmark for a person with impaired sight? Which modalities are used to encode landmarks in the memory and how does this relate to cognitive load? What contributions to wayfinding are made by senses other than vision and does language play a role?

7.1.2 Wayfinding for B&PS

Navigation difficulties have been identified as one of the main hindrances to independent living for people who are B&PS (Golledge, 1993). Yet those with impaired vision are still able to integrate spatial information (Ungar, 2000) such that they can encode a mental map, recall a route, give directions and point out short-cuts (e.g. Rieser, Guth, & Hill, 1986; Passini & Proulx, 1988). Rieser et al. (1980) showed that when estimating distances between landmarks, B performed as well as S for functional (route-based) distance. However, CB were significantly less good than LB or S participants at estimating Euclidean distances (as the crow flies). Ungar (1996) found almost all estimates made by visually impaired participants were closer to functional than Euclidean distance. When Bigelow (1991) asked children to point to objects along a familiar route, those with congenital blindness invariably pointed along the route they would take to reach a landmark rather than pointing towards the landmark itself.

Brambring (1982) compared descriptions given by B and S along a specific route. Blind participants reported less about their environment but their descriptions gave more specific information about distance and body rotation and were generally more detailed. Passini and Proulx (1988) showed that B and S were almost equally good at creating a map of an indoor route, 250 meters long, that they had been guided along twice and completed once alone. When asked to verbalise their journey, blind participants cited a greater variety of "small features, such as a radiator, a doorframe, or an ashtray, that went unnoticed by the sighted wayfinder" (p. 243).

These studies suggest that what sighted people traditionally conceive as landmarks (i.e. large-scale, visually distinctive objects at key turning points) need to be reconsidered in the context of wayfinding without sight. In O&M terminology, a "*landmark"* is defined as "any familiar object, sound, odor, temperature, tactile, or visual clue that is easily recognized, is constant, and has a discrete permanent location in the environment that is known to the traveler" (Hill & Ponder, 1976, cited in Blasch et al., 1997, p. 42). Landmarks are contrasted with '*information points*' that are clues to location provided by two or more sensory stimuli. However, this may be a false distinction given that a landmark is also likely to be perceived through more than one modality. The long cane, for example, is a navigation aid that provides both tactile and auditory information and both are necessary to the user for accurate identification of objects at key points on a journey. Schenkman (1986) demonstrated that ground surfaces encountered with a cane when either auditory or haptic-tactile information was deliberately suppressed were often identified incorrectly.

Blasch et al. (1997) distinguish between "*primary landmarks"* that are constant, and "*secondary landmarks"* that maybe intermittent or easily missed. However, in understanding differences in navigation strategies according to acuity of sight, it may be more useful to distinguish between landmarks of any size that allow people to orient themselves in a particular direction and "*milestones"* that are encountered along the pathway between landmarks i.e. objects that, in the absence of a vista, assist in path integration. In topological terms, landmarks are found at nodes along the journey. Milestones are found along the links between the nodes. Both landmarks and milestones can be divided into proximal or distal depending on whether or not they lie within reach, as illustrated in Fig. 7.1.

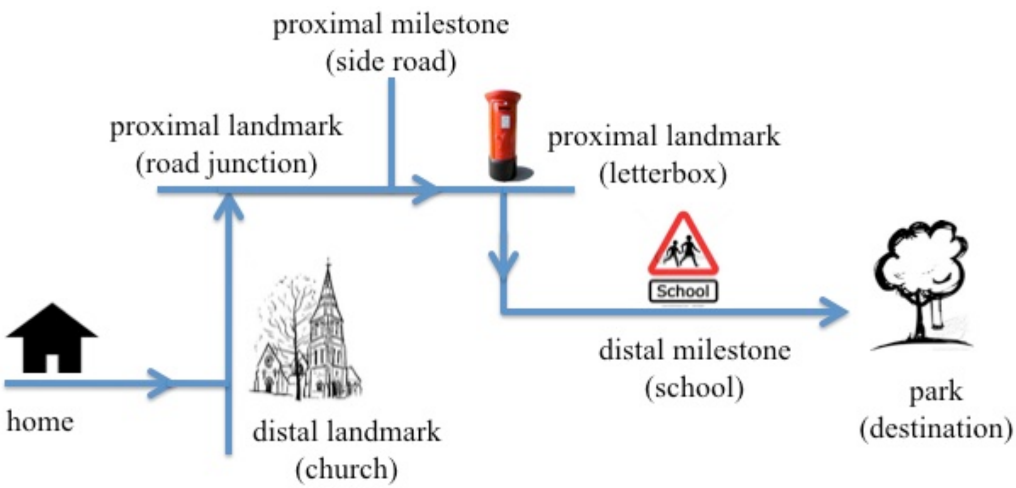


Fig. 7.1 Categorisation of landmarks and milestones (landmarks denote turning points; milestones punctuate the pathway)

7.1.3 Comparing navigation strategies

Empirical studies designed to compare the effectiveness of wayfinding strategies often rely on laboratory situations that do not translate well to the real world (Kitchin & Blades, 2002). Real environments tend to provide more multisensory feedback that may be particularly helpful for B&PS. The studies cited above compared S and B participants undertaking the same route. Building on Brambring's research design, the current study asked 9 sighted participants and 9 B&PS participants to describe a journey with which they were familiar. This allowed individuals to recall a route they knew extremely well. The expectation was that they could recount such a route more fully thus allowing a comparison to be made as to the type and number of navigational cues encoded in the memory. To aid comparability and usefulness, participants were asked to recall a journey made on foot. According to the RNIB, walking is the most frequently used method of locomotion by people whose vision is impaired (Worsfold & Chandler, 2010). Walking binds other journey elements together but is currently that part of a journey “where there is the least amount of information or assistance to help the individual” (2010, p.8). As one of the interviewees in Chapter 3 put it: “Walking if you're totally blind, and even if you're going somewhere you know, is a chore – it’s an act of total concentration 100% of the time” (ID005).

Previous studies have concentrated on what features of a route B and S people identify. This study was explicitly concerned with *how* such features are identified through a range of multisensory cues, and, more specifically, which modalities are encoded in the memory. For example, the auditory compensation hypothesis would suggest that B&PS would rely mostly on auditory cues. However, given the evidence presented so far it is possible B&PS will rely as much on other modalities, or on language. In line with previous research, it was expected that participants would report more small-scale features of their route compared with sighted participants. In particular they would report more milestones along the pathway. Landmarks and milestones were likely to be close to hand and identifiable by modalities other than vision.

7.2 Method

7.2.1 Participants

9 volunteers with impaired sight took part with no payment for their participation. All were known to the researcher and were selected to cover a range of sight conditions (see Table 7.1). All had been registered blind for at least 10 years, were experienced long cane users and judged to be skilled, independent travellers. All had completed a school education and 8/9 had a college degree.

ID	Gender (Age)	Sight status	Visual acuity*	Aetiology	Mobility aid
B&PS1	F (31)	PS	4	Retinitis pigmentosa	Long cane/ wheelchair user
B&PS2	M (31)	EB	1	Persistent hyperplastic primary vitreous	Long cane
B&PS3	M (31)	PS	4	Optic atrophy	Long cane
B&PS4	F (44)	LB	2	Shrunken eye syndrome	Long cane
B&PS5	M (47)	LB	2	Cone rod dystrophy, photophobia, cataracts	Long cane
B&PS6	M (59)	CB	1	Retinopathy of prematurity	Long cane
B&PS7	M (63)	CB	1	Retrolental fibroplasia	Long cane
B&PS8	M (73)	LB	1	Retinitis pigmentosa	Long cane
B&PS9	F (75)	PS	3	Macular degeneration	Long cane

*Visual acuity coding (after Douglas et al., 2006). Sight status: CB = congenital blindness; EB = early blindness; LB = late blindness; PS = partially sighted.

Table 7.1 Blind and partially sighted participants

9 sighted volunteers, also not paid to participate, were recruited of comparable age, gender and educational background. This resulted in the following sample (N = 18, F = 6): B&PS = 9; S = 9; aged 31 – 75 years, m = 50.06, SD = 16.91.

7.2.2 Procedure

Interviewed individually, each participant was asked to imagine undertaking a familiar journey and to describe the route. Descriptions were recorded using Audio Engineering's Field Recorder i-phone application, FiRe!, and transcribed by the researcher. A clear protocol for route terminology was determined and is outlined below. Route descriptions were analysed for quantitative and qualitative differences in landmarks and milestones reported and the modality through which they were identified. Information on distance was also noted. B&PS participants were also given the chance to talk about wayfinding in general. These additional comments inform the discussion section.

7.2.3 Route Terminology

7.2.3.1 Route Complexity

Participants were given no limitation on the length or type of route they described. The only stipulation was that it should be "familiar". It was therefore possible that some journeys would be longer than others. Route complexity was determined by the number of junctions/turnings specified rather than distance travelled.

7.2.3.2 Landmarks

Landmarks were defined as objects that served an orienting purpose at, or immediately prior to, a change of direction e.g. “at the pub I turn right” (B&PS1). “Distal” landmarks lay beyond the subject’s immediate peripersonal space e.g. “by the white house on the corner” (S4). “Proximal” landmarks lay within reach and could be touched either directly by hand or felt underfoot, or indirectly by long cane e.g. “at the end of the hedge, there’s a tree and I walk round the tree” (B&PS9).

7.2.3.3 Milestones

Milestones were defined as objects along the pathway where no change of direction was required. As with landmarks, milestones were divided into distal e.g. “I go past some industrial buildings” (S1) and proximal e.g. “on your left is a container for Metro papers” (B&PS5). There was an additional category of obstacles i.e. milestones to be avoided e.g. “you walk across the forecourt - avoiding bollards and lamp posts” (B&PS2).

7.2.3.4 Distances

Metrics were defined as “*objective”* e.g. metres/yards, “*body-based”* e.g. paces/steps, or “*environmental*” meaning items encountered along the route e.g. “I count 4 dips in the pavement and those are dips for car drives where cars can drive in” (B&PS8). Distances were “*precise*” e.g. “exactly 100 steps” (B&PS7), or “*vague*” e.g. “after about 10 yards” (S7).

7.2.3.5 Modality

Modality of landmark or milestone identification was either explicitly stated e.g. "you might hear the children out playing" (S9) or inferred. Vision, for example, was taken to be the relevant modality when a landmark could be identified only by appearance e.g. "a small, flint cottage in the shape of a church" (S6). Information pick-up was deemed multimodal if more than one modality was specified, or if no specific modality was stated e.g. for sighted participants, it was assumed that a busy road could be heard and seen.

7.3 Results

The number of words used by participants ranged from 151 to 994. However one participant (B&PS1) used 2,690 words. Z-scores showed this description to be an outlier ($z = 3.66$). The participant was excluded from the quantitative analysis, but retained for illustrating qualitative differences between the two sight groups (B&PS/S). The total sample was small and the data was not normally distributed, so quantitative differences were compared using the non-parametric Mann-Whitney U test, and the median is the average used throughout. Most participants described their journey in the first person; 4 participants (S = 3; B&PS = 1) used "you" as though they were describing the route for someone else to follow. This difference between groups was not significant.

7.3.1 Route complexity

There was no significant difference between sight groups in the number of junctions/turnings specified: B&PS (mdn = 2); S (mdn = 2) $U = 23.5$, $z = -1.24$, n.s., r = -.30. This suggests the routes described were of similar complexity. However, the B&PS group provided significantly longer descriptions (mdn = 461 words) compared with the S group (mdn = 212 words) *U* = 12, z = -2.31, p = .021, r = -0.56.

7.3.2 Landmarks, milestones and obstacles

There was no difference between sight groups in the total number of landmarks reported: B&PS (mdn = 8.5); S (mdn = 10) *U* = 26.5, z = -.92, *n.s.,* r = 0.22. For the B&PS group, landmarks were significantly more likely to be proximal: B&PS (mdn = 5); S (mdn = 3) *U* = 17.5, z = -1.80, p = .037, r = 0.22. The B&PS group also reported a significantly greater number of proximal milestones (mdn = 14.5) compared with the S group (mdn = 6) *U* = 14, z = -2.14, p = .036, r = -.52. They also mentioned obstacles whereas the S group did not: B&PS (mdn = 3.5); S (mdn = 0) *U* = 13, z = -2.30, p = .027, r = -0.56. However, the S group reported a significantly greater number of distal milestones (mdn = 7) compared with their B&PS peers (mdn = 4.5) *U* = 14, z = -2.14, p = .036, r = -0.52.

7.3.3 Modality

The modalities in which landmarks and milestones were reported, as a percentage of the total, are shown in Table 7.2. Unsurprisingly the S group were more likely to identify waymarkers by sight: visual landmarks: S (mdn = 4); B&PS (mdn = 0) *U* = .00, z = -3.67, p < .001, r = -.89; visual milestones: S (mdn = 10); B&PS (mdn = 0) *U* = .00, z = -3.66, p < .001, r = -.89.

	Landmarks		**Milestones**	
	B&PS	S	B&PS	S
Visual	0%	82%	1%	80%
Auditory	2%	2%	11%	1%
Olfactory	0%	0%	2%	0%
Body-based	16%	0%	28%	2%
Multimodal	82%	16%	55%	17%
Absence	0%	0%	3%	0%
Total	100%	100%	100%	100%

Table 7.2 Percentage of landmarks and milestones in each modality

By contrast, the B&PS group identified more waymarkers using a combination of two or more modalities such as sound and touch, sound and smell, or sound, touch and body-based information: multimodal landmarks B&PS (mdn = 5.5); S (mdn = 1) $U = 11$, z = -2.48, p = .015, r = .60; multimodal milestones B&PS (mdn = 9.5); S (mdn = 2) $U = 7.5$, z = -2.77, p = .004, r = -.67. Absence of milestones was also noted by the B&PS group (mdn = 3) but never by the S group: $U = 16.5$, z = -2.10, p = .05, r = .50 (see Figs. 7.2 and 7.3).

7.3.4 Metrics

Compared with their sighted peers, the B&PS group made more references to metrics (B&PS total = 61; S = 45). These were significantly more likely to be "specific" rather than "vague": B&PS (mdn = 5.5); S (mdn = 3) $U = 15.5$, z = -1.99, p = .046, r = -.48 (see Fig. 7.4). There was no between-groups difference in the frequency of objective (e.g. metres/yards) nor environmental metrics (e.g. flowerpots or side roads), but B&PS participants used significantly more body-based metrics (e.g. paces): B&PS (mdn = 1); S (mdn = 0) $U = 16.5$, $z = -2.05$, $p = .05$, $r = -.50$.

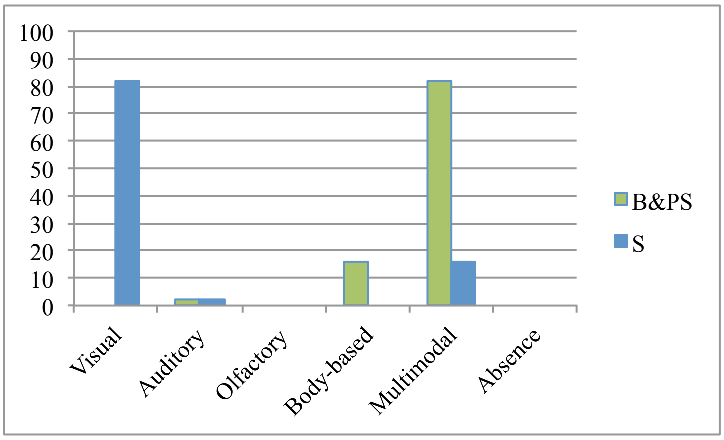


Fig. 7.2 Modality of landmarks (percentage) by sight group

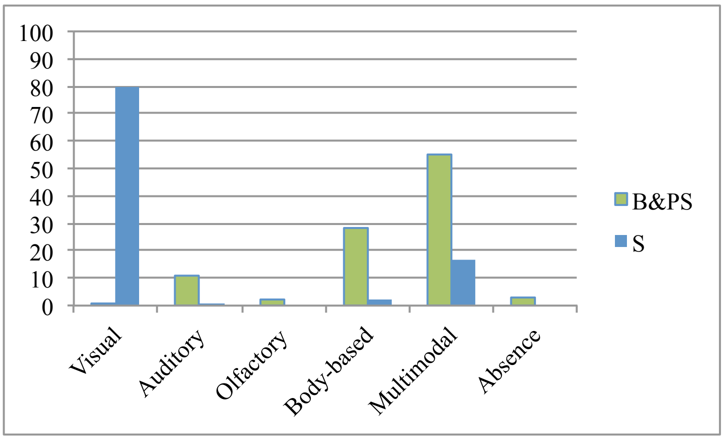


Fig. 7.3 Modality of milestones (percentage) by sight group

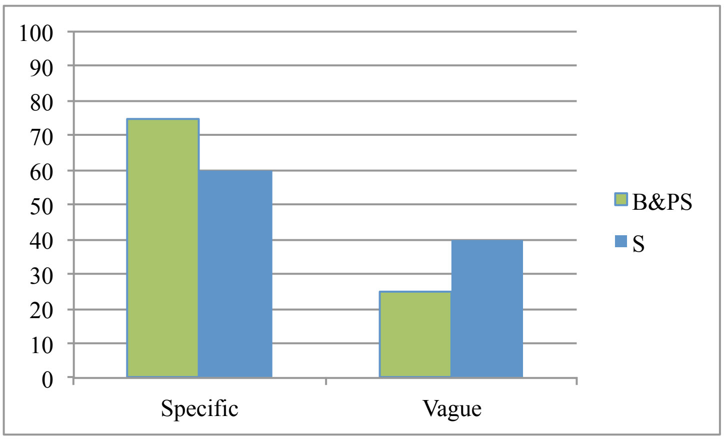


Fig. 7.4 Percentage of metrics that are “specific” rather than “vague”

7.4 Discussion

This study analysed verbal descriptions of familiar routes walked by B&PS and sighted people in order to explore differences in navigational strategies. Routes encompassed both rural and urban environments. By asking people to describe their routes, the aim was to elicit what type of spatial information is gathered and held in an individual's event-related knowledge base.

Participants' journeys were of similar complexity, judged by mentioning a similar number of junctions and a similar number of landmarks along the route. However, in line with previous studies (e.g. Brambring, 1982; Passini & Proulx, 1988), the B&PS group used over twice as many words in their descriptions as the S group. It is possible that people with impaired vision are simply more verbose. However, it is more likely the difference in word length reflects a greater level of detail. B&PS participants reported significantly more milestones than their sighted peers. They also included more references to distance and these were specified with greater precision. It is likely that B&PS people are more used to verbalising their journey, either as a deliberate encoding strategy, or because, when negotiating a route for the first time, a sighted guide or an O&M trainer has described the route aloud (Martinsen et al., 2007). For those with sight, orienting towards a vista allows much of the information along the pathway to remain implicit. However, landmarks need to be pointed out at a turning point i.e. the moment when one uninterrupted vista is exchanged for another.

7.4.1 Modality of information pick-up

Confounding the assumption that vision is necessary for landmark navigation, both groups referred to similar numbers of distal landmarks. However, the B&PS identified significantly more proximal landmarks and relied on significantly different modalities. Sighted participants referred specifically to visual information pick-up for 82% of landmarks e.g. "an untidy front garden" (S8) or "a white house" (S3). 2% of landmarks were identified by audition alone. The remaining 16% were identified by multimodal means either explicitly, e.g. with visual clues corroborated in sound: "The school is on the corner, so you might hear the children out playing" (S9), or implicitly e.g. by visual information which could also be perceived through other, body-based information: "You go down a dip" (S3). The high percentage of specifically visual information was strikingly similar for both landmarks (82%) and milestones (80%).

The B&PS group relied on multimodal sources for 82% of landmarks. As the majority of these were proximal, this may be explained by the use of the long cane. The cane provides auditory and tactile feedback, together with proprioceptive information through the angle of the wrist or arm (Sauerburger & Bourquin, 2010). As one man reported:

> "So then I'll check with my white cane and 'plonk' if I hear the railings then I know where I am. Once the railings stop and my white cane goes round the corner… I know I'm at the pavement" (B&PS8).

Like vision, the cane provides information in advance, for instance preparing the user for a change of surface underfoot:

> "Just inside after about a metre is a piece of carpet … and I tap my cane on that and the texture is obviously different from the pavement and from the tiled floor of the shop so that confirms where I am" (B&PS8)

As well as receiving multimodal information from a single source, B&PS often cited more than one landmark at the same location. As one PS participant reported:

> "The next landmark I notice is another drain on the ground and I know that at that point I've got to turn slightly right and also I can see that it's lighter because there's no building there." (B&PS1)

Perhaps surprisingly, purely auditory information was used as rarely by B&PS participants as by their sighted peers, again for only 2% of landmarks. This may be because sound can be considered both helpful and a hindrance. For one participant, the increasing volume of traffic noise indicated she was approaching a busy road:

> "…so gradually the traffic gets louder and when it gets louder I slow down [because] I'm looking for the dropped kerb on my right." (B&PS1)

However, B&PS participants confirmed the suggestion from Chapter 6 that sound is hard to isolate. A loud sound source, for example, can obscure more subtle auditory cues. As one man put it: “My obstacle sense is slightly impaired by the roar of traffic to my right” (B&PS7). By contrast, quieter cars can also be a problem in judging when to cross the road:

> “In years gone by I would then stand and listen to traffic. When it was quiet and I couldn't hear any vehicles coming then I’d cross. These days with Prius cars and the like which are very quiet and quite quick I tend to stand there and wait for the kindness or generosity of strangers.” (B&PS8)

Silent obstacles can obscure auditory information from the surrounding environment. A wall, for example, will reflect the sound of a tapping cane. However, as another participant explained: “a parked car …blocks sound, so then you can't hear where the wall is when you cross over” (B&PS4). A further explanation for auditory information rarely appearing as a unimodal source may be because it is bound to other modalities. For instance, sound may be triggered by an action, such as reverberating footsteps or the strike of a cane, which may simultaneously also provide body-based or tactile cues.

Unlike S, B&PS people showed a wider disparity between preferred modalities for landmarks and milestones. In particular, unimodal sources were cited for 18% of landmarks but 45% of milestones. It is possible that detection of milestones is deemed to be less critical than that of landmarks. As almost twice as many milestones were reported than landmarks, there was a greater likelihood that if one was missed, another could compensate. 11% of milestones were detected through audition alone (compared with 2% of landmarks) and 2% of milestones were detected by olfactory information (compared with 0% of landmarks). These were references to shops identifiable by smell:

> "I quite often know roughly where we are on that street because there's a butcher's and I can smell it and there's a Subway and I can smell that as well." (B&PS1)

As with sound, smell was deemed by some participants to be unreliable. One woman commented:

> "Smells can be quite deceptive because sometimes you can have one smell and the next day…it could be covered by rain, it could be covered by somebody chopping that tree down or something." (B&PS4)

Nonetheless, olfactory information featured in three B&PS participants' route descriptions (B&PS 1; B&PS2; B&PS7).

By contrast, smell was never mentioned by S participants. This does not necessarily mean olfactory cues were not perceived. Zhou et al. (2010) carried out a laboratory experiment exploring binocular rivalry, presenting different images to each eye. They found that the presence of a smell, congruent to one image and not the other, lengthened the time the congruent image was visible, but in a manner that was "automatic, essentially independent of cognitive control and partly subconscious" (2010, p. 1356). The current study suggests that while

B&PS people consciously encode cues from non-visual modalities, sighted people do so only very infrequently, if at all.

7.4.2 Body-based sensory information

Waller and Greenauer (2007) found that, for S, wheeling participants along a route made little difference to the accuracy of their spatial memories. They speculated that “the transient nature of body-based information may render it particularly unsuitable for storing an enduring mental representation” (2007, p. 331). This study casts doubt on that suggestion. B&PS participants mentioned the use of body-based information when identifying 16% of landmarks and 28% of milestones, in particular reporting small changes of level, such as negotiating front door steps or stepping up onto the kerb. Three blind participants (B&PS2; B&PS4; B&PS5) also referred to orientation of their bodies in relation to fixed points such as walls or kerbs, or to changing the angle in which they were travelling: e.g. “I angle myself inwards towards the shop-line” (B&PS2). There were no references to speed of movement from the S group, but two of the B&PS participants commented on change of pace along certain sections of their route e.g. “I then reach a lamp-post conveniently sited in the middle of the pavement. This I generally slow down for and keep to the inside of” (B&PS5). Rather than body-based information being unsuitable to store in long-term memory, it is perhaps more readily accessed when external sensory information is absent or unreliable. Alternatively, as with “external” modalities such as sound and smell, body-based information may be actively encoded by B&PS people, but only subconsciously registered by people with sight.

7.4.3 Types of landmarks, milestones and obstacles

Roads were mentioned as landmarks and milestones by both sight groups, either generically e.g. “the main road” (S1) or by name e.g. “Geoffrey Road” (B&PS5). Both groups, therefore, made use of semantic information. While sighted people might look for a sign to identify a road, for blind people this was not an option. However, knowledge of a road name was still deemed useful. One participant explained that it allowed route-details to be checked with a passer-by if necessary (B&PS4). Presumably the name of a road also provides a useful tag for accumulated multisensory information. Jones (2007) points out the ability of a word to conjure up a particular place, citing Gerrig (1993) who describes the many connotations of the name “Texas” and Glenberg (1997) who does the same for “Amazon”.

Proximal landmarks and milestones that were mentioned by both groups included pedestrian crossings, gates or flights of steps. In line with other studies, however, for the B&PS group, proximal waymarkers also included smaller items of street furniture such as rubbish bins or lamp-posts and continuous elements, known as “*shorelines”* (e.g. Blasch et al., 1997), including kerbs, walls, hedges or railings that could be followed for a particular stretch of the route. Although providing useful reassurance that they were on the right path, 75% (6/8) B&PS participants referred to street furniture as obstacles to be avoided:

> “There are a lot of things in the way including posts, poles and various other bits of street impedimenta but when I have negotiated all those with the stick and with my ears I carry on.” (B&PS7)

Moveable elements were especially problematic: e.g.

> "In spring and summer there are restaurants which have tables outside the restaurant as well as metal A-boards and God knows what else…And so in spring and summer after about 9 or 10 double steps, I move to the right a little bit." (B&PS8)

Atmospheric and temporal conditions not only alter the location of milestones but also the ability of B&PS participants to identify them in a number of modalities: e.g.

> "I can detect objects by the change in the sound if the conditions are good and it's not very windy and it's not, say, raining" (B&PS2)

> "It was very grey when I went out… the colours were just grey, brown, a bit of green but it was dark green so they began to merge together and I found it quite difficult to distinguish between the hedge and the ground which rather alarmed me." (B&PS9)

This would suggest that navigation strategies are adjusted not just according to the demands of the route, but according to the weather, ambient light conditions, season or time of day.

There was no mention of obstacles by sighted people presumably because sight enables a traveller to steer easily around them without consciously registering their presence. Certainly mobile robots can effectively detect obstacles from visual appearance alone by constantly comparing the colour of image pixels from the obstacle with the colour of pixels from the ground (Ulrich & Nourbaksh, 2000). Sighted people may not commit moveable items to memory precisely because they are not always in the same place. However for B&PS it is important to remember stretches of route where unpredictable objects might be encountered, in order to limit the chance of collision. It is likely that smaller-scale objects of the type relied on by those with impaired sight are more susceptible to change than, for example, a building or a geographical feature. Again, the large number of milestones mentioned by the B&PS group may reflect the fact that some might not be encountered every time that journey is undertaken.

Researchers recognise that, for S, routes are easily disrupted if a landmark or other information is removed (Bennett, 1996; O'Keefe & Nadel, 1978). This study confirmed that, for B&PS, an absence of milestones/obstacles can be especially problematic: e.g.

> "the pavement slopes up to the right and there's grass …which is also a bit confusing. If you land on that grass you could end up skittering about for quite a bit before you get back onto your proper path." (B&PS6)

7.4.4 Metrics

Foo et al. (2005) argue that for S, routes are best remembered as a topological network, with connectivity and ordinal relations taking preference over precise geometric relationships. They liken the resulting mental model to “a cartographic map printed on a rubber sheet” (p.196). In this study, S participants made equal use of specific and vague metrics, on average mentioning 3 of each type, and in general referring to much longer distances: “I follow this track for about ¾ mile” (S6). As it is likely that a track would offer plenty of milestones over such a distance, in many different modalities, it seems safe to conclude that for S, metrics, however vague, are used instead of milestones. Time is another vague metric commonly used by sighted people (e.g. I walk for about 5 minutes) although it was not mentioned by any of this particular sample. As vision allows any obstacles en route to be negotiated easily, some general indication of distance or duration until the next significant landmark is all that is required.

The B&PS participants made significantly more references to metrics overall, and these were more likely to be specific rather than vague. While S used metrics in preference to milestones, B&PS used metrics for stretches of journey precisely where no milestones where available:

> “I am absolutely at the door of the bank after I've counted 100 steps. I have got a clue at 30 steps, a zebra crossing with a butcher's shop that I can smell and then the last 70 steps I have to count.” (B&PS7)

In the absence of environmental information, then, B&PS individuals resort to abstract methods of dead-reckoning. Although this would appear to be at considerable cognitive cost, one LB participant suggested it became easier with practice:

> "I count [paces]…and if I'm interrupted by somebody who knows me who stops to say hello, somehow my mind now clamps onto the last number in my head so I can just pick it up." (B&PS8)

Although B&PS sometimes reported vague metrics, these tended to refer to very small distances e.g. "I go slightly to the left" (B&PS1) whereas sighted people were vague about long stretches of route: "when I get to the end after probably 500 metres, I then turn right" (S1). In addition, the B&PS group was more likely to report body-based metrics. Although these may seem vague compared with objective measurements, they could be specified with great precision: e.g. "it is normally about 25 paces …if I'm wearing boots it'll be 23 or 24" (B&PS8).

7.5 Limitations

This was a small-scale study using participants with diverse sight characteristics describing a wide variety of routes. Participants were free to interpret the task as they wished. There was no specific instruction to give as much detail as possible. However, B&PS people clearly described much more information about their journeys than their sighted peers. Assigning modality of landmarks and milestones was problematic given the potential multimodal nature of much of the information. All cues reported by sighted participants were regarded as multimodal unless they comprised information, such as the colour of a building, which could only be accessed by vision. Similarly all cues reported by B&PS participants were assumed to be multimodal where no modality was specified. This was the case for all those transmitted by use of a cane. While some of the B&PS group specifically referred to this as multimodal, others did not, perhaps assuming knowledge on the part of the researcher. However, there is also the possibility that as the multimodal information provided by the cane was perceived directly, it was less likely to be verbally recalled. This would seem to support the hypothesis put forward in Chapter 6 that proximal information results from implicit perceptual associations, whereas distal information is explicitly formed and therefore more readily reported.

7.6 Summary of Study 5

This study was designed to explore differences in the way one aspect of real-world spatial information, that relevant to navigation, is encoded and recalled by S and B&PS people. While S consciously register information in few modalities other than vision, B&PS rely on non-visual modalities, occasionally unisensorily but more often in combination. In the absence of a vista, orienting towards a landmark and judging progress towards it is accomplished by the use of multiple small-scale milestones, accessed by non-visual means. Metrics are specified where milestones are not available. In verbal descriptions given by sighted people, multisensory information is rarely made explicit, milestones are overlooked and metrics are vague and sketchy for long stretches of the route. By contrast, verbal descriptions given by B&PS people are highly detailed. This supports the hypothesis of implicit/explicit differences in encoding according to sight characteristics when the route is committed to memory. This real-life experience may, in turn, help to explain why sighted people need perceptual, non-verbal sound to activate multimodal imagery in their SSM, while it provides less benefit to B&PS.

Although the contribution of auditory information to navigation, and by extension spatial mental models, was only one aspect of this study, comments by B&PS participants supported the idea that non-verbal sound can be unreliable and that multiple sound sources can hinder rather than help. Other research with B&PS has shown that verbal information is regarded as more useful than non-verbal information in navigational settings. For example, research for the Talking Images project (RNIB, 2003), which explored the experience of B&PS in galleries and museums, found that verbal directions of the kind included on an audio guide were considered more helpful than tactile maps, with 66% (137/208) of B&PS respondents reporting tactile maps to be of no benefit. Golledge et al. (2004) also found that B&PS preferred speech input to guide them. More recently Neuville and Trassoudaine (2009) showed that words were considered preferable for wayfinding compared to a directional, non-verbal audio signal.

These findings question a study by Klatzky et al. (2006) in which participants in conditions of high cognitive load (completing an *N*-back task) took less time to complete a path when guided by virtual sound (i.e. artificially-generated, rather than arising from the natural environment) than when guided by spatial language. This prompted the researchers to suggest virtual sound would be beneficial to B&PS. However, their participants were blindfolded sighted people. As suggested in Chapter 2, it is possible that, by processing speech in the visual cortex, B&PS people reduce the cognitive load generated by explicit directions. Moreover, verbal information is likely to be a deliberate part of their regular encoding strategy. This familiarity with reliance on verbal mediation may help expand the capacity of WM compared with S (Rokem & Ahissar, 2009).

Studies 3, 4, and 5 all point to the importance of verbal information for B&PS people, both in the real world and in the mediated environment. However, the impact of visual impairment on presence has yet to be explored within an AV medium. This is addressed in the final study.

Chapter 8 Study 6: Words or Sounds? The Impact of Verbal Information (AD) on Emotional Elicitation and Presence

8.1 Summary

For B&PS people, words have so far been shown to be as effective as non-verbal sounds in evoking mental representations and in making mediated environments feel immersive and natural. It has been argued that this stems from real-life experience in which verbal information assists everyday actions such as navigation. As participant ID011 put it "I'm so used to people talking to me, whether it's my watch, my computer, whatever…that I just take it as read". The final study extends the exploration of presence from an audio-only medium (Study 4) to an audiovisual one. As outlined in Chapter 2, the sighted audience engages with the source material using audio and visual cues combined. For B&PS audiences, the visual information is conveyed verbally through the addition of AD. The question that arises is the extent to which that visual information is necessary? One criticism levelled at AD is that it provides details that may be detected from the soundtrack alone. This stems from the assumption that SFX are more evocative than words and that B&PS have heightened auditory perception. Although the findings from the studies presented so far have questioned both those assumptions, the implications of masking music and SFX with speech have yet to be tested. It is possible that providing semantic information reduces emotional engagement. As a Study 1 participant put it (cited in Chapter 3):

> "Sometimes in plays you can really get sucked into the atmosphere you know and the feelings that are being built up, and sometimes another person speaking [i.e. the audio describer] can distract from that and I prefer to sacrifice a bit of information for being immersed in the atmosphere more." (ID015)

For sighted people, more emotive material has been linked to higher ratings of presence (e.g. Dillon, 2006). This final Study aims to measure the impact of AD in an audiovisual medium on emotion elicitation; on the four dimensions of presence in the ITC-SOPI, and also on Social presence (affect and empathy). The studies so far have relied on self-report. This study is designed to combine self-report with more "objective" physiological measures, comparing explicit and implicit feedback on AD user experience.

8.2 Introduction

8.2.1 Visual information, non-verbal auditory information and emotion

Kreifelts, Wildgruber and Ethofer (2012, p. 225) state that "when judging their social counterpart's emotional state, humans predominantly rely on non-verbal signals". The majority of these are visual, such as facial expressions, gestures and posture, but also include non-verbal vocalisations and prosody i.e. the rhythm, emphasis and intonational aspects of speech. In an fMRI study Kreifelts and colleagues (2007) showed that, for S, a combination of auditory and visual information led to higher success in classifying emotions than either the auditory or visual component alone. Furthermore, emotional signals in one modality also affect processing of emotional signals in the other. For example, Maiworm et al.

(2012) showed that the ventriloquism effect was weakened when preceded by a threatening auditory stimulus (syllables spoken in a fearful voice). For B&PS, does the loss of visual reinforcement for auditory cues lead to less effective emotional processing? The auditory compensation hypothesis would suggest not. However, under the linguistic compensation model, emotional processing may benefit from explicit (verbal) reinforcement.

8.2.2 The impact of visual impairment on accessing emotion cues

As discussed in Chapter 4, a high proportion of CB children exhibit social interaction problems that show similarities with autism (Perez-Pereira & Conti-Ramsden, 2005). One aspect is a delayed ability to develop Theory of Mind i.e. to understand another person's point of view (Minter et al., 1998). This is thought to result from lack of access to visually-presented emotional cues such as gesture and facial expression (Hobson, 1993). Indeed Hobson and Bishop (2003, p. 342) suggest "vision has a special role in linking children with other people and with others' attitudes towards a shared world". Lacking the bimodal advantage of sighted children, blind children are also significantly worse at identifying vocally-expressed emotions (Pring, 2002). Yet Pring also suggests the majority of those who demonstrate an early deficit are able to "catch up", with verbal ability a key factor in their success.

It is clear from the interviews in Chapter 3 that when it comes to mediated environments, B&PS participants can pick up affective auditory cues unassisted, but they are also aware that this is only one part of the equation. Accessing visual cues to emotion ranked top of one CB participant's list of what he wanted from

AD i.e. “the manners, the gestures and the looks of the person” (ID004). If language helps blind children access non-verbal social interactions, it seems reasonable to infer that this extends to all B&PS people. As one LB participant pointed out in Chapter 3: “If I made a lot of money I’d have an audio describer with me at all times, telling me facial expressions or…that person doesn’t get your joke, or whatever” (ID007).

8.2.3 Emotion, imagery and presence

In Eardley and Pring’s (2006) study of mental imagery (see Chapter 5), emotion was shown to be an important encoding factor. They suggest it explains why B retrieved as many autobiographic memories as S when prompted by words high in “visual” imagery. “Sunset”, for example, was linked by one B participant not to a visual image of pink-streaked clouds but to an occasion shared with her husband on a boat at twilight, bringing to mind the excitement audible in his voice, as well as other, non-visual, perceptual experiences such as the sound of the rippling water and the feel of the air on her skin.

Recall of embodied perceptual experience makes it easier to create a first-person perspective and thus a PERF within which to experience a mediated environment. As discussed in Chapters 2 and 6, the creation of a PERF has been directly linked with presence (e.g. Vogeley & Fink, 2003; Wirth et al., 2007). This suggests that more emotive content might increase ratings of presence in B&PS people. A positive correlation between emotive content and presence has certainly been demonstrated in S (Baños et al., 2004; Freeman et al., 2005). Dillon (2006), for example, showed that video clips from emotionally arousing films generated higher levels of presence than neutral footage.

As discussed in Chapter 2, empathy and affect also enhance Social presence and, by extension, enjoyment (Shedlosky-Shoemaker et al., 2011) both in the mediated environment and in the real world. B&PS participants in Study 1 spontaneously mentioned wanting to be able to share emotional responses, such as laughter, with fellow members of the audience. The benefit is not limited to positive emotions: audiences can also derive enjoyment from sadness or fear (Oliver, 1993; Tamborini & Schiff, 1987; Zillmann, 2003). For example, bursts of emotion-laden visual stimuli such as gloomy lighting, an alarmed facial expression, or the shadow of a villain looming moments before the character himself appears (Smith, 1999) create what film critic Thomas Sutcliffe calls “the delicious interlude between suspicion and confirmation” (2000, p.115). The auditory equivalents include sounds arising from the action (e.g. heavy breathing, or footsteps); SFX (both environmental sounds e.g. the hooting of an owl and non-naturalistic sounds e.g. an amplified heartbeat) and music that underscores a scene. Such auditory prompts can lead the audience to interpret the visuals according to a particular schema. In a horror movie, for example, scary music leads us to expect that a door, slowly opening, will swing wide to reveal a murderer rather than a person collecting on behalf of a charity.

8.2.4 Emotion, presence and AD

Emotion differs from other aspects of presence in that there is no need to suspend disbelief (Gerrig, 1993). We experience the emotion directly. As Jones (2007, p.

121) points out, “our awareness of the falsity of an emotional stimulus does not stem the tide of the emotion itself”. While watching a scary movie, sighted people often respond naturalistically, hiding their eyes or experiencing the fight/flight reflex as if the events in the film were real. Sad movies are nicknamed “weepies” with good reason: Our eyes prick with tears as we watch events unfold onscreen. A “behavioural realism” approach to presence proposes that the more present we are in the mediated environment, the more we exhibit such behaviours (Freeman et al., 2000).

Study 4 showed that for B&PS people, words can be as effective as non-verbal sound for creating a sense of presence but does that extend to experiencing emotional content as real? There is a difference between recognising an emotion and experiencing it. All participants (16/16) in Study 1 referred to occasions, in the cinema or a theatre, when they knew something funny was happening but were not able to access sufficient information that they could laugh themselves. Comments of two PS participants suggest that AD helps compensate for lack of direct visual information: ID006 and ID015 reported physical responses (a sharp intake of breath and closing their eyes) when watching gory scenes with AD. Comments from one LB participant (ID010) supports the suggestion that AD also helps by creating the emotional context:

> ID010: “ …If you’re seeing an Agatha Christie for instance, and they’re sitting round the fire or whatever and the door opens very quietly and somebody comes in, but those people sitting in the room haven’t heard it, but the AD tells you that “so and so has come into the room and he’s

> moving towards the people by the fire" and you know that something is going to happen then [it] does help enormously… and if you hadn't known that somebody was coming in and that something was about to happen it would come to you almost as a shock but also it would lose its interest, because by knowing something's going to happen, which a sighted person can see, you can build it up inside you…it does add a terrific amount for me."

Despite such qualitative evidence, and as discussed in previous chapters, some have argued that the interaction between the visuals and soundtrack leaves the B&PS audience (and the audio describer) in an impossible bind. While compensating for information carried by the visual stream, the AD reduces access to the information carried in auditory stream by masking with speech the non-verbal auditory cues (such as music and SFX) that themselves contribute to emotional and contextual understanding. Too much AD risks adding unnecessarily to the cognitive load; too little can fail to reduce the load that stems from having to work out what is happening from only partial information.

8.2.5 Prosody, AD delivery and emotion

Sander et al. (2005) point out that while a large amount of research has focussed on processing visual emotion cues, particularly faces, relatively little study has been made of emotional signals in the voice. Emotional information is carried, therefore, not just in the words of the dialogue but in the actors' delivery, including pitch, pace, emphasis and inflection. Similarly AD is received not as a written text, but as a spoken one. It therefore also consists of more than just semantic information. To what extent might the AD audience infer emotional meaning from the content of the description, and to what extent from the affective prosody of the describer's delivery?

Ofcom guidelines suggest that voice of the describer should be "unobtrusive and neutral, but not lifeless or monotonous and the delivery should be in keeping with the nature of the programme" (ITC, 2001, p. 10). Belin et al. (2000) have shown that, for sighted people at least, emotionally salient auditory information attracts bottom-up, exogenous attention in a way that a neutral auditory stimulus does not. Certainly participants in Study 1 commented that AD worked better when the delivery of the describer matched the emotional cadence of the scene. As ID015 put it: "I find it good when it's obvious there's an emotional atmosphere going on that [the audio describers] speak quieter, I like that". Arguably, a delivery style that is neutral or incongruent with the emotional content of the scene could reduce presence.

8.2.6 Electronically-generated speech

B&PS people are very familiar with "neutral" voices in the form of electronically-generated speech. A report for the RNIB (Cryer & Home, 2009) shows that it is commonly used to access computers, when operating gadgets such as a mobile phone, and for leisure reading or study. Their respondents suggested that the use of text-to-speech (TTS) software was acceptable for accessing informational content, such as newspaper items, but less so for fictional content such as e-books. Interestingly, while some respondents felt electronic speech lacked "human" aspects of interpretation and emphasis, others suggested one benefit of neutral delivery was that the listener was free to add their own, imaginative, interpretation.

A Polish study explored the use of TTS as a cheaper alternative to Human Voice (HV) AD (Szarkowska & Jankowska, 2011). The study was based on the reception of a multilingual film (Pedro Almodóvar's *Volver*), with the dialogue captioned in Polish, so in addition to a standard AD, the describer (in this case an electronic voice) also spoke the captions aloud, providing "audio subtitles". The researchers found that although half the B&PS participants stated a preference for HV AD, the majority were willing to accept electronically-generated speech if it led to more source material being audio described. One drawback of the study was that participants were not able to make a direct comparison with HV AD. Moreover, the effect of TTS on emotional engagement, empathy and presence was not measured.

8.2.7 AD and presence: A previous study

As mentioned in Chapter 2, a previous study (Fryer & Freeman, 2012a; 2012b) used the closing scenes of David Lean's 1945 classic *Brief Encounter* to explore the effect of HV AD on presence. Specifically it compared No AD with two types of HV AD: a standard description and a "cinematic" AD that added extra visual information, giving details of camera shots and editing. This style was specially developed on the premise that sighted people report fast-cut film sequences to be more arousing (Kraft, 1986). The study compared sight groups (no, some, full useable vision) on the 4 dimensions of presence defined by Lessiter et al. (2001) in their ITC-SOPI (see Chapter 5) i.e. Spatial presence, Engagement, Ecological validity and Negative effects. For B&PS participants, AD generally enhanced presence, rather than detracted from it. However, differences emerged according to sight characteristics. With No AD, levels of Spatial presence and Ecological validity were significantly lower for PS participants than B participants. This was explained as an effect of everyday experience. The grainy black and white footage of the film made it difficult for PS participants to decode any useful visual information. For B participants, used to relying purely on sound, the film's atmospheric soundtrack seemed naturalistic and was sufficiently informative for the action to be clearly located.

S reported lower levels when the stimulus was audio described than when watching the stimulus alone. By contrast, B&PS participants' presence ratings were not only higher with AD than without, but also higher with AD than ratings reported by S watching with No AD. Waterworth & Waterworth (2003b) found that, for sighted people, presence ratings were higher for media forms that elicited predominantly concrete (perceptual) processing, compared with forms that depended on more abstract (semantic) processing, so these results may reflect a difference in the relationship between perceptual and semantic processing between S and B&PS. Alternatively, in the absence of (reliable) vision, B&PS people may have benefitted from the ability to shape their own mental representation from the verbal description. As with Nunez & Blake's naïve users of flight simulator games (2006) referred to in Chapter 2, B&PS users may have been unhindered by flaws in the visual content, enjoying higher imaginal presence.

Cinematic (but not standard) AD led to significantly higher ratings of Spatial presence for all B&PS participants. It was proposed that the use of cinematic terms giving the camera point-of-view added a greater sense of movement and depth, as characters approached the viewer or moved away, or walked from one side of the shot to another. As seen in Chapter 7, spatial mental models can be created by modalities other than vision and mental maps can be encoded through language. This cinematic information also significantly increased Ecological validity and Engagement ratings but only for PS. B participants reported little difference in Engagement between the two AD styles. It is possible that this masked a difference between CB and LB: cinematic terms may have placed a greater cognitive load on CB although the study did not test this.

Another limitation of the study was the poor quality of the visual stimulus. This may have reduced presence levels of sighted participants. There was also no explicit attempt to investigate the impact of AD on empathy or emotion elicitation.

8.2.8 AD, emotion and presence: Hypotheses

This study addresses those issues using a selection of movie clips, assembled by Gross and Levenson (1995), known to elicit specific target emotions in sighted people. If AD obscures emotional cues from the soundtrack (SFX, music and dialogue) then those target emotions should be elicited more strongly with no AD than with AD. This is in line with the auditory compensation model and leads to the following hypothesis:

H1: For levels of the target emotion: No AD > AD

Alternatively, according to the linguistic compensation model, affective visual information (e.g. facial expression) replicated in words will enhance the emotional impact. This gives rise to:

H2: For levels of the target emotion: No AD < AD

To separate the semantic content of the AD from the affective content of its delivery, the same AD script was trialled in two conditions: HV and TTS.
In line with the findings of Belin et al. (2000), the benefit of paralinguistic emotional information contained in the delivery of the human voice leads to the third hypothesis:

H3: For levels of the target emotion: TTS AD < HV AD.

If the findings of Study 5 translate from an audio-only to an audiovisual medium, reducing access to SFX by the addition of AD should have no adverse effect on Spatial presence, Ecological validity or Engagement. Furthermore, should H2 prove correct, all three positive presence ratings should increase in line with strength of emotional response. This leads to:

H4: For levels of presence (positive subscales): No AD < AD

If H3 proves correct such that presence levels increase in line with the strength of the target emotion, then:

H5: For levels of presence (positive subscales): No AD < TTS AD < HV AD.

8.3 Method

8.3.1 Participants

Participants in this study received £10 for taking part and were recruited through the RNIB's social media channels and personal contacts. This resulted in the following sample: N = 19 (Male = 10), aged 24 – 77 years (m = 48.53 years, SD = 13.64). All participants were registered blind or partially sighted and demonstrated a range of sight characteristics: CB/EB = 4; LB = 11 (4 = no vision; 7 = light perception); PS = 4 (see appendix C1 for details).

8.3.2 Materials

Gross and Levenson's original corpus was put together in 1995. Dillon (2006) showed that, a decade on, not all of these clips were still effective in eliciting their target emotion in sighted people. Drawing on Dillon's revised corpus, 3 target emotions were initially selected: Amusement, Fear and Sadness. However, AD has to be woven around dialogue. Potential clips in the Amusement category (e.g. the orgasm scene from *When Harry Met Sally*) had no suitable gaps for AD. It was therefore decided to focus on Fear and Sadness. Clips from three films (*Bambi*; *Truly, Madly, Deeply*; and *Four Weddings and a Funeral*) were selected from the Sadness category. Dillon suggested only two clips for retention in the Fear category: *The Shining* and *Silence of the Lambs*. These were supplemented by the opening scene of a more recent horror film, *The Ring*. This was of similar duration to the longest Sadness clip and judged to be suitably scary. The clips are listed in Table 8.1.

Fear	**Scene**	**Duration**	**Details**
Silence of the Lambs (**Lambs**)	Basement chase	3m 28s	1991, Director: Jonathan Demmer, BFI classification: 18
The Ring (**Ring**)	Opening scene	5m 31s	2002, Director: Gore Verbinski, BFI classification: 15
The Shining (**Shining**)	Boy playing in hallway	2m 11s	1980, Director: Stanley Kubrick, BFI classification: 15
Sadness			
Bambi (**Bambi**)	Death of Bambi's mother	3m 7s	1942, Disney animation BFI classification: U
Truly, Madly, Deeply (**Truly**)	Woman in Counselling	3m 14s	1991, Director: Anthony Minghella BFI classification: PG
Four Weddings and a Funeral (**Weddings**)	Funeral Speech	5m 29s	1994, Director: Mike Newell BFI classification: 15

Table 8.1 Stimuli for Emotion Study

An AD script for each clip was written by the researcher who is a professional audio describer. The scripts can be found in Appendix C3. The researcher recorded the human voice (HV AD) version for each clip. The same AD scripts were "voiced" using the Text-to-Speech feature in the Mac Operating System 0SX 10.8 (Mountain Lion). The English voice "Emily" was chosen to match the voice of the researcher as closely as possible in accent and timbre. Together with the non-described versions of the film clips, this resulted in a total of 18 stimuli: i.e. 3 styles (No AD/ HV AD/ TTS AD) x 3 clips x 2 categories (Fear/Sadness).

8.3.3 Design

Each participant watched 6 clips – one from each film. The order and condition of presentation was randomised across participants with the proviso that 2 clips (one from each emotion category) were presented with no AD, 2 with HV AD and 2 with TTS AD.

All the supporting documentation for this study is shown in Appendix C, including samples of the written consent form; the question sheet, personal details sheet and debriefing. All questions were read aloud by the researcher and required either a verbal or haptic response (see *Measurement procedure* below).

8.3.4 Measures: Self-Report

8.3.4.1 Elicited Emotion Scale

Gross and Levenson (1995) created a 16-item Elicited Emotion Scale (EES). For each emotion, participants rate the greatest amount of it they experienced at any time during the clip using a 0 – 8 scale. 0 means “not feeling even the slightest bit of the emotion” and 8 is “the most you have ever felt in your life”. Of the original 16 items, 6 relevant to this study were used: Arousal; Interest; Fear; Sadness; Tension; Confusion. Although Fear and Sadness are the target emotions, Tension was included as, in Dillon’s study, it was rated as highly as Fear for one clip: Lambs. Arousal and Interest were there to assess general engagement that was not emotion-specific. This would allow comparisons to be made between clips across the 2 emotion categories. Confusion was included as it may arise where lack of access to visual information (in the No AD condition) or possibly the AD itself has a negative effect on comprehension (for example using "visual" terms that CB participants might not understand). Given the emphasis of AD guidelines on helping AD users "follow the plot" (Rai et al., 2010), Confusion could provide useful insights on the links between comprehension, emotional engagement and presence. Participants were also asked whether they had previously watched the film from which the clip was taken, and, if yes, whether with or without AD.

8.3.4.2 ITC-SOPI short form

To avoid fatigue, the ITC-Sopi (Lessiter et al., 2001) was presented in its short form. This consists of three items for each of the positive presence subscales: Sense of Physical Space, Engagement, and Ecological Validity. A single, composite question addressed Negative effects: "I experienced sensations such as dizziness, disorientation, nausea, a headache, eye-strain or tiredness". The ITC-SOPI uses a 5-point Likert scale (1 = strongly disagree; 5 = strongly agree).

8.3.4.3 Identification and Empathy

Plantinga & Smith (1999) have shown that identification with characters and empathy with characters are determinants of emotional response to narrative film. 2 simple questions addressed these aspects, taken from Dillon's study (2006):

1. How much did you identify with the concerns of one or more of the characters; that is to what extent did the clip address issues relevant to you?
2. How much did you empathise with the characters; that is feel the same emotions they were experiencing?

These questions used a 7-point Likert scale (1- not at all; 7 – very strongly)

8.3.4.4 Semantic understanding

In order to separate emotional engagement from semantic understanding of emotion, participants were asked a question that explicitly addressed this:
What emotion do you think the clip was trying to make you feel?

8.3.4.5 Measurement procedure

Sighted people are able to complete self-report questionnaires without making an explicit verbal judgement. Apart from online questionnaires, this is not usually the case for B&PS participants. In order to avoid this problem and to help deal with potential confusion arising from different scales for each measure, a haptic device (a "slider") was especially created for this study (see Fig. 8.1). This was a wooden board, slightly larger than A4 size, made in two sections (a top-side and an under-side) with a narrow gap for an answer sheet to be slid between. The paper was printed with horizontal lines and positioned by the researcher so that each line in turn lay beneath a wooden pointer towards one end of the board. In response to a verbal question from the researcher, participants slid the pointer to one side or the other i.e. towards the positive or negative end of the scale. The final position was marked by the researcher. Before each question, the slider was located in its central position by the participant, adjusted by the researcher if necessary. The next question or statement was read aloud and the participant was able silently to move the slider to indicate their response. Participants had a chance to familiarise themselves with the haptic slider before the session began. The slider was used for all measures that required a Likert scale.

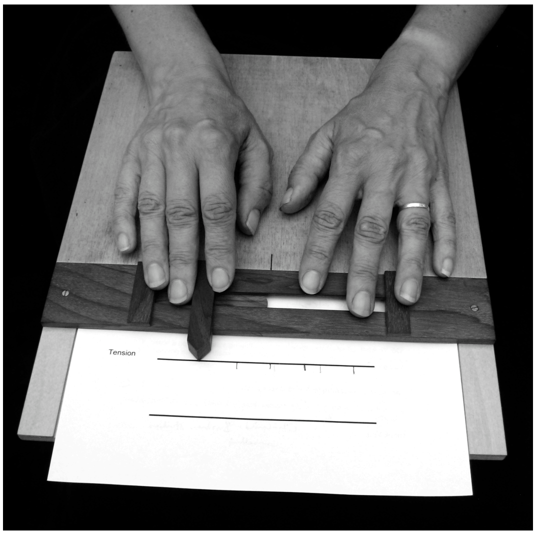


Fig. 8.1 The haptic slider.

8.3.5 Measures: Physiological

The emotions we experience are linked to physiological arousal (Levenson, 2003). As described in Chapter 2, physiological responses are triggered by the Central Autonomic Network (CAN) via the autonomic nervous system. The CAN coordinates exogenous and endogenous stimuli, allowing an individual to adjust their physiological, behavioural and cognitive response in accordance with emotional context, resulting in appropriate emotional expression (Hagemann, Waldstein & Thayer, 2003). In addition to the self-report measures, therefore, two implicit indicators were used to record physiological reactions: Electrodermal activity (EDA) and Heart Rate (HR). In previous studies both of these have been shown to differ according to emotional content (Wilson & Sasse, 2000; Mandryk et al., 2005; Dillon, 2006). Data was collected using a custom-made polygraph[7]. Data was transmitted to a PC with custom-made software used to collect and display it. The same software automatically triggered each film clip that was displayed on a separate screen. Physiological data was recorded continuously and time-stamped with the onset time of each stimulus. An inbuilt delay of 6 minutes between clips allowed sufficient opportunity for the self-report data to be collected, leaving a "resting" window (minimum 30s) for HR and EDA levels to settle before the next clip began.

8.3.5.1 Heart rate

HR was monitored by electrocardiography (ECG) using Biotrace disposable AG/AgCI electrodes with snap connectors. These were attached to the back of the right and left hands and to the right leg just above the ankle. The electrodes were connected by leads to the data acquisition amplifier, which was in turn connected to the PC.

8.3.5.2 Heart Rate Variability

In addition to HR, Heart Rate Variability (HRV) was calculated, using root mean square deviations (RMSSD) of the HR time series to assess the variance from normal-to-normal inter-beat intervals (i.e. the oscillation of intervals between heartbeats). HRV is suppressed when an individual is under stress or engaged in mental effort. However, it may increase in conditions of cognitive overload (Mandryk et al., 2005).

---

[7] Software and hardware was designed and made by Rob Davis, Dept. of Psychology, Goldsmiths College

8.3.5.3 Electrodermal activity

EDA was measured with electrodes attached to the distal phalanges (fingertips) of the index and middle finger of the right hand. Skin resistance was sampled at 200Hz and measured in kilohms.

8.3.6 Procedure

Participants took part individually. Where a participant attended with a sighted companion, the companion was invited to wait at a nearby café. Two participants brought a guide dog with them. In each case, the dog was released from its harness and curled up quietly for the duration of the session. Each session lasted for approximately 90 minutes. On arrival, the participant was settled into a chair facing a screen a short distance away. The consent form was completed. The participant was introduced to the haptic slider and completed the EES to provide a baseline. The electrodes were attached and connected to the polygraph. When the participant was ready, instructions were read aloud (see C2.3) and, after a 30s pause, the first clip was played. The clip ended with a blank screen and silence for 30s, after which the researcher handed the haptic slider to the participant and all the self-report measures were completed. At this point the participant was informed how much time would elapse before the next clip began. They were instructed to sit still and to take a moment to clear their mind of all thoughts, feelings and memories. The procedure was repeated for each clip. When all 6 clips had been viewed, the electrodes were removed. The questions on the Personal Details sheet were read aloud and the researcher recorded the responses. Participants were also invited to ask any questions about the clips or make any general comments. Finally, a selection of participants was asked to recall which of the clips they had watched with No AD, which with TTS AD and which with HV AD. After a debriefing the participant was reunited with their dog or sighted companion, or accompanied to the underground station by the researcher.

8.4 Results: Self-report data

8.4.1 Comments on the analysis

As the two emotion categories (Fear/Sadness) were expected to yield different results (e.g. ratings of Fear were likely to be higher than ratings of Sadness for the Fear clips, and lower than Sadness for the Sadness clips), data for each category was analysed separately. There was large individual variation indicated by high SDs. Consequently, a within-subjects design was chosen using Bonferroni corrections for multiple comparisons. Scores on the EES were calculated as an increase/decrease from baseline. As all participants showed a high degree of Interest (4 and above) and most (14/19 and 13/19 respectively) showed a high degree of Arousal and Tension (4 and above) before beginning the task, the clip ratings for these emotions are correspondingly low. In one session, the fire alarm rang while the participant was viewing his final clip (TTS AD: Fear). His data has been excluded in comparisons of Fear clips, but retained for the Sadness clips which had all been viewed before the alarm.

A MANOVA for EES ratings showed that, in all conditions, which clip was viewed and whether participants had seen the clip before (either with or without AD) had no significant effect on emotion ratings.

8.4.2 Subjective responses: Fear

8.4.2.1 Elicited Emotion Scale

Mean EES scores for the three conditions (No AD; TTS AD; HV AD) are shown in Table 8.2. Paired samples t-tests for emotion ratings in the No AD condition showed that Fear was elicited significantly no more strongly than any other emotion except for Interest: t (17) = 3.05, p = .007. This was the same in the TTS AD condition: Interest: t (17) = 2.86, p = .011. By contrast, in the HV AD, Fear was elicited significantly more strongly than Arousal: t (17) = 2.09, p = .05; Sadness: t (17) = 3.58, p = .002; Interest: t (17) = 3.78, p = .001 and Confusion: t (17) = 2.50, p = .022. In the HV condition, only the difference between Fear and Tension was not significant: t (17) = 1.40, p = .178.

A within-participants repeated measures ANOVA showed a main effect of condition on three emotions: Fear: F (2, 34) = 4.21, p = .024; Arousal: F (1.34, 22.8) = 3.61, p = .038; and Confusion: F (2, 34) = 5.16, p = .014. Mauchly's Test indicated that the assumption of sphericity had been violated for Arousal (χ^2 (2) = 10.802, p < .001) so a Greenhouse-Geisser correction was used.

Planned paired comparisons between conditions showed that HV AD (but not TTS AD) gave rise to significantly higher ratings of Fear than No AD (mean difference = 1.78, p = .016). HV AD (but not TTS AD) was also significantly more arousing: mean difference = 1.61, p = .039. Both TTS AD and HV AD were significantly less confusing than No AD: TTS AD/No AD mean difference = -1.94, p = .027; HV AD/No AD mean difference = -2.06, p = .003 (See Fig. 8.2).

	No AD (fear)	TTS AD (fear)	HV AD (fear)
Arousal	0.33 (3.61)	0.94 (1.86)	1.94 (1.76)
Interest	-1.22 (3.64)	-0.61 (2.17)	-0.33 (2.95)
Confusion	3.17 (2.64)	1.22 (3.26)	1.11 (3.22)
Fear	1.67 (3.66)	2.00 (3.40)	3.44 (2.45)
Tension	1.89 (3.07)	2.22 (2.78)	2.28 (2.63)
Sadness	-0.17 (3.40)	-0.11 (2.87)	0.11 (3.03)

Table 8.2 Means (SD) for increase/decrease in EES ratings from baseline for Fear clips.

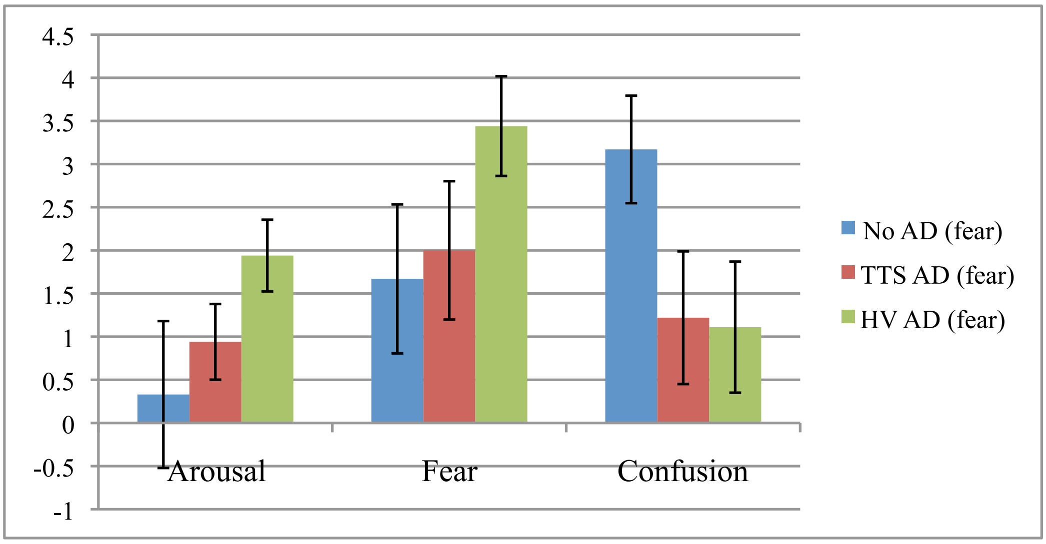


Fig. 8.2 Mean (SEM) increase/decrease in EES ratings from baseline for 3 emotion states: Arousal, Fear and Confusion.

8.4.2.2 Empathy and Identification

Mean scores for Empathy and Identification with one or more characters are shown in Fig. 8.3. There was no main effect of condition on either variable, although there was a trend towards ratings of Empathy being higher with HV AD: t (17) = -1.88, p = .076.

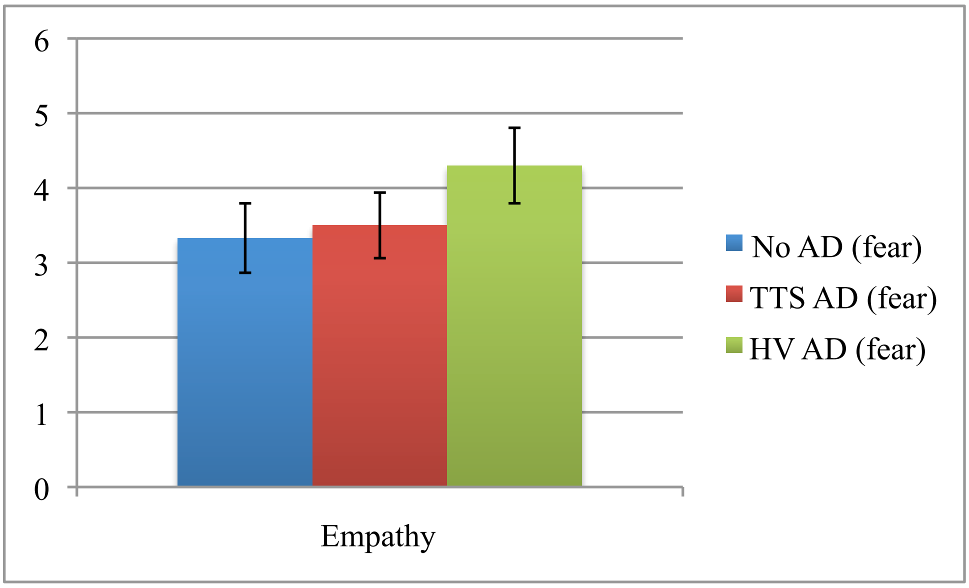


Fig. 8.3 Means (SEM) of Empathy and Identification induced by Fear clips in each condition

8.4.2.3 Presence scales

Mean scores (SDs) for the 4 Presence scales are shown in Fig. 8.4. A one-way repeated measures ANOVA showed a main effect of condition (No AD/ TTS AD/ HV AD) on all dimensions: Spatial presence: F (2, 34) = 3.48, p = .042; Engagement: F (2, 34) = 6.60, p = .004; Ecological validity: F (2, 34) = 4.42, p = .019; and Negative Effects: F (2, 34) = 3.44, p = .044). Adjusting for multiple comparisons, the main effect on Spatial presence was no longer significant (p = .079). However, HV AD was more significantly more engaging than either No AD or TTS AD (mean difference between HV AD/TTS AD *and* HV AD/No AD = .87, p = .02). For Ecological validity, HV AD (but not TTS AD) was significantly more "natural" than watching with No AD (mean difference = .87, p = .002) and participants reported significantly fewer Negative effects with HV AD (mean difference = -.611, p = .035).

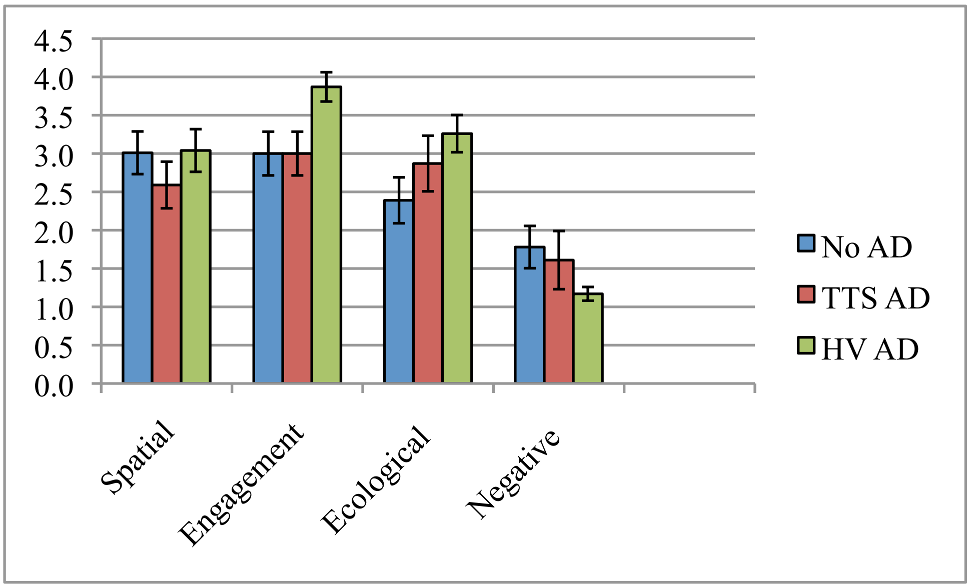


Fig. 8.4 Means (SEM) of presence ratings induced by Fear clips in each condition

8.4.2.4. Emotional Understanding: Accuracy

In answer to the question “What emotion do you think the clip was trying to make you feel?” participants scored 2 points for an accurate answer (i.e. “Fear” or “Scary”); 1 point for a related emotion (e.g. “Tension”); 0 points if they were unable to name an emotion or named an unexpected one (e.g. “Sadness”). A within-subjects repeated measures ANOVA showed no difference in score across conditions (F (2, 34) = 1.71, p = .196). All participants asked to recall in which condition they had viewed each clip were able to do so without error, except for one who incorrectly recalled a Fear clip viewed with TTS as having been viewed with No AD (E14).

8.4.2.5 Fear clips: Summary

For the Fear clips, then, there was no evidence to support H1: that B&PS people could more successfully access the target emotion in the No AD condition. Only in the HV AD condition did Fear (together with Tension) stand out from other possible emotion states. There was partial support for H2, that B&PS participants accessed the target emotion to a greater degree with AD than without. This was only true for the HV rather than the TTS condition, although both HV and TTS AD were deemed significantly less confusing than No AD. H3, therefore, was fully upheld. The prosody in the describer's voice (rather than the semantic meaning of the words) was able to elicit significantly higher levels of the target emotion (Fear). HV AD also led to significantly higher ratings of Arousal. Although participants reported higher ratings for how strongly they identified or empathised with the on-screen characters with HV AD, the increase was not significant. As for presence, the mediated scenes were significantly more engaging, felt more natural and there were significantly fewer Negative effects only with HV AD. There was also a non-significant trend towards greater Spatial presence with HV AD. H4 was therefore partially upheld and H5 fully upheld but only for 2 dimensions of presence, Engagement and Ecological validity.

8.4.3 Subjective responses: Sadness

8.4.3.1 Elicited Emotion Scale and Sadness clips

There was a very different pattern for the Sadness clips. Mean scores for the three conditions (No AD; TTS AD; HV AD) are shown in Fig 8.5. Paired samples t-tests showed that the clips successfully elicited the target emotion (Sadness) significantly more strongly than any other emotion in all three conditions: No AD: Arousal: t (18) = 3.25, p = .004; Interest: t (18) = 4.97, p < .001; Fear: t (18) = 5.08, p < .001; Confusion: t (18) = 3.76, p = .001; Tension: t (18) = 3.89, p = .001; TTS AD: Arousal: t (18) = 2.38, p = .029; Interest: t (18) = 4.31, p < .001; Fear: t (18) = 3.37, p = .003; Confusion: t (18) = 2.92, p = .009; Tension: t (18) = 3.09, p = .006; HV AD: Arousal: t (18) = 3.11, p = .006; Interest: t (18) = 5.05, p < .001; Fear: t (18) = 6.07, p = .001; Confusion: t (18) = 4.66, p < .001; Tension: t (18) = 4.20, p = .001.

A one-way, within participants repeated measures ANOVA comparing conditions showed a main effect only on Confusion: F (1,18) = 5.23, p = .01. Planned paired comparisons showed that HV AD (but not TTS AD) was significantly less confusing than No AD: mean difference = -2.05, p = .008.

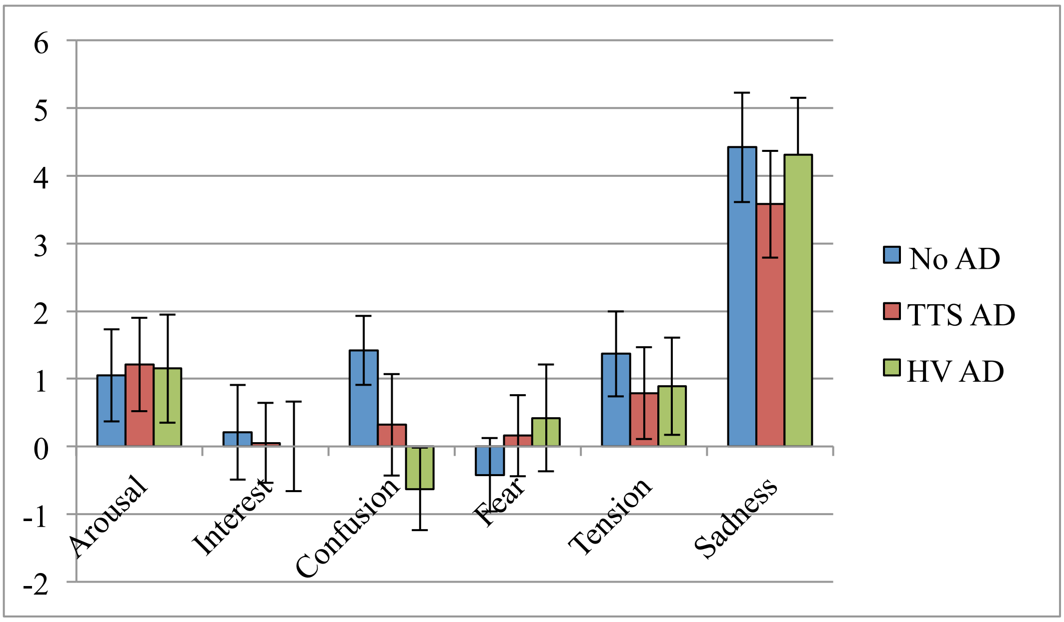


Fig 8.5 Mean (SEM) increase/decrease in EES ratings from baseline for Sadness clips

8.4.3.2 Empathy and Identification with Sadness clips

Means scores (SDs) for Empathy or Identification with the characters are shown in Fig. 8.6. Although the overall levels are higher than for the Fear clips, the Sadness clips also showed no main effect of condition (No/TTS/HV AD).

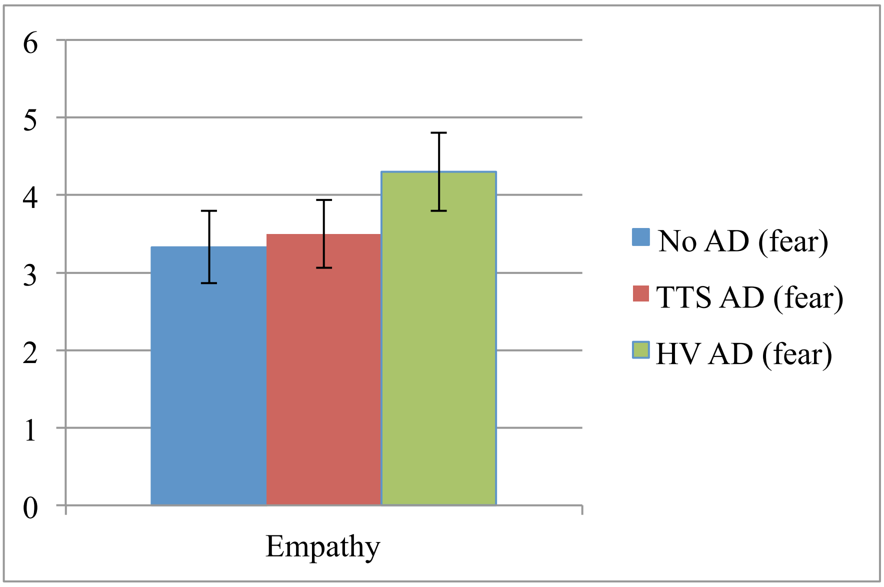


Fig. 8.6 Means (SEM) of Empathy and Identification induced by Sadness clips in each condition

8.4.3.3 Presence

Means scores (SE) for the 4 Presence scales are shown in Fig. 8.7. A one-way repeated measures ANOVA showed no main effect of condition (No AD/ TTS AD/ HV AD) on any dimension of presence.

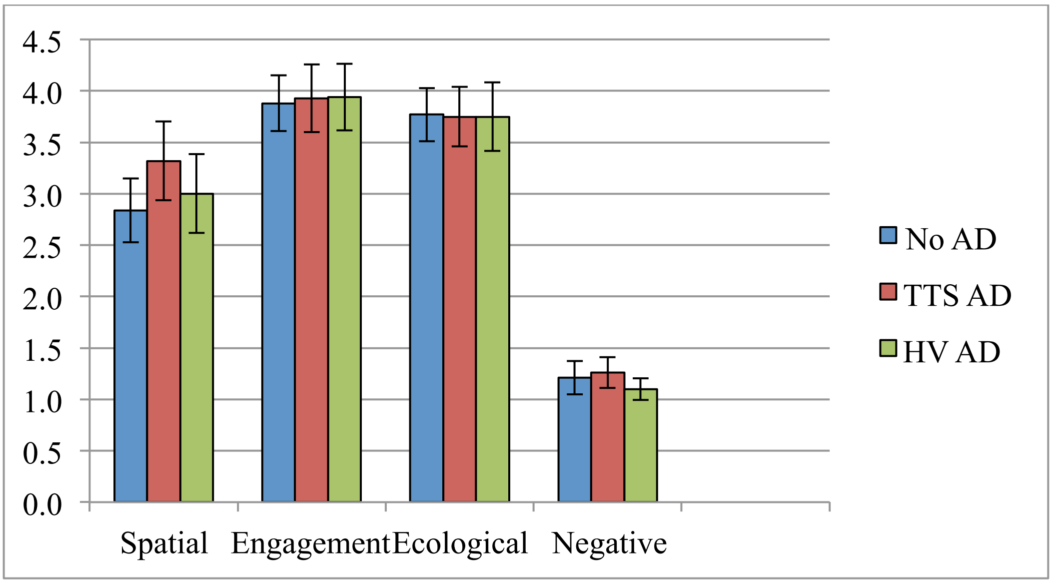


Fig. 8.7 Means (SEM) of presence ratings induced by Sadness clips in each condition

8.4.3.4 Emotional understanding: Accuracy

In answer to the question "What emotion do you think the clip was trying to make you feel?" participants scored 2 points for an accurate answer (i.e. "Sadness" or "Grief"); 1 point for a related emotion (e.g. "Loss"); 0 points if they were unable to name an emotion or named an unexpected one (e.g. "Anger"). There was no within-subjects difference in scores across conditions: F (2, 36) = 2.37, p = .107. There was no correlation between accuracy and familiarity for any of the conditions.

8.4.3.5 Sadness clips: Summary

Overall, there was little variation between conditions. For Sadness clips, HV AD significantly improved understanding. However, this did not impact on level of elicited emotion or sense of presence, Empathy or Identification. Results for the Sadness clips uphold none of the hypotheses. Differences between the two emotion categories (Fear/Sadness) are examined in more detail below.

8.4.4 Correlations between presence, emotions, Identification and Empathy

One aim of this study was to see whether a more intense emotional experience correlated with higher ratings of presence and, in particular, how AD might disrupt or enhance that. Tables for all correlations are shown in Appendix C6. Significant correlations are reported below. A summary of correlations between presence measures and target emotion is presented in Table 8.3.

Is there a significant correlation between…	**No AD (fear)**	**TTS AD (fear)**	**HV AD (fear)**	**No AD (sad)**	**TTS AD (sad)**	**HV AD (sad)**
… Spatial Presence and Engagement	No	No	Yes*	Yes*	Yes**	Yes**
… Ecological validity and Engagement?	Yes*	No	Yes*	Yes**	Yes**	Yes**
… Spatial Presence and Ecological validity?	Yes*	Yes**	Yes**	Yes*	Yes**	Yes**
… target emotion and Spatial presence?	No	No	No	No	No	No
… target emotion and Engagement?	Yes**	No	No	Yes**	No	Yes*
… target emotion and Ecological validity?	No	Yes*	No	Yes**	No	No
… target emotion and Negative effects?	No	Yes**	No	No	No	No

*. Correlation is significant at the 0.05 level (2-tailed).
**. Correlation is significant at the 0.01 level (2-tailed).

Table 8.3 Summary of correlations between target emotion and presence scales (significant correlations shaded in grey)

As an example, Fig. 8.8 shows the correlations between the positive dimensions of presence for the Fear clips with HV AD:

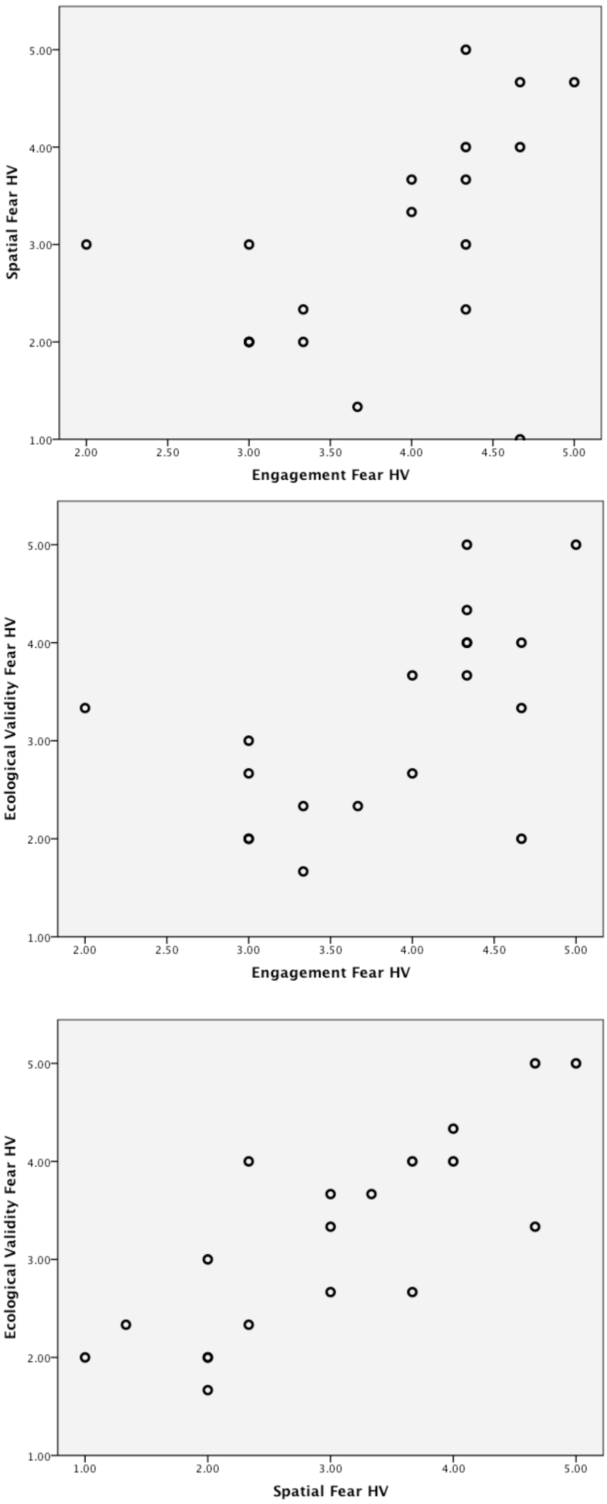


Fig. 8.8 Correlations between the positive dimensions of presence for Fear clips with HV AD

Fig. 8.9 shows the correlations between target emotion (Fear) and Engagement for each condition (No AD, TTS AD, HV AD). A summary of correlations between presence measures and affect (Identification and Empathy) is presented in Table 8.4.

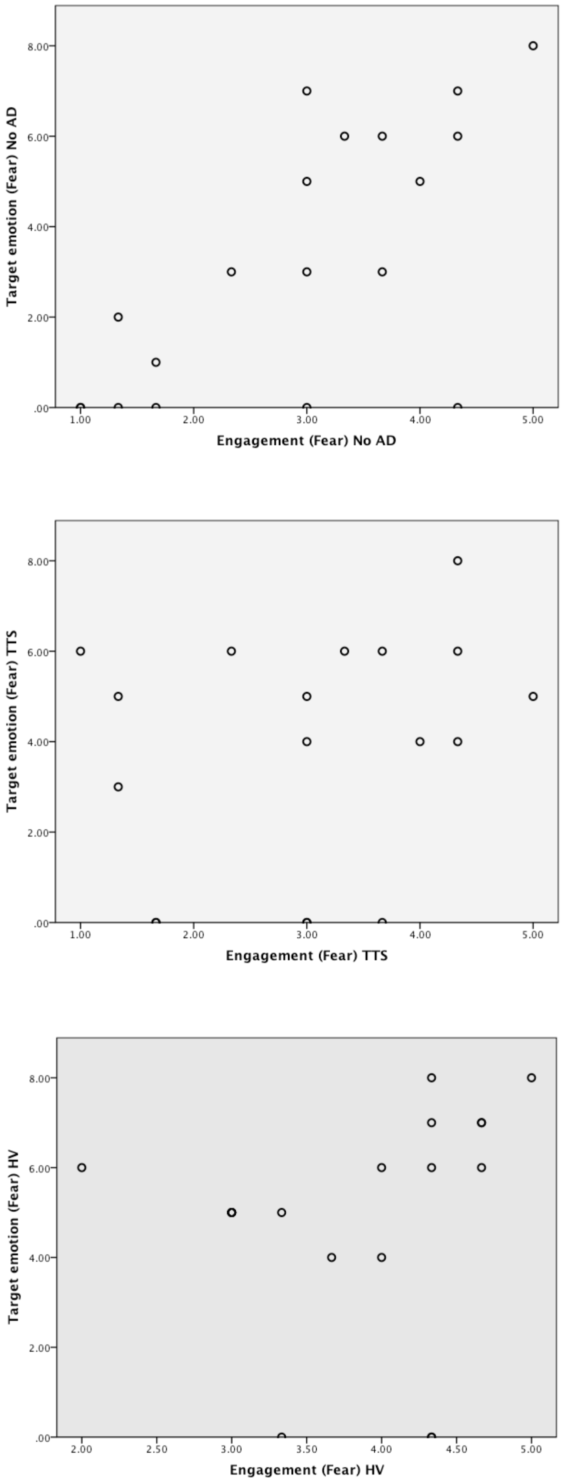


Fig. 8.9 Correlations between target emotion (Fear) and Engagement

8.4.4.1 No AD: Fear

Bivariate correlations showed that, in the No AD condition, the target emotion (Fear) was significantly positively associated with Engagement (r = .66, p = .002). Engagement was also significantly associated with Tension (r = .62, p = .005); Arousal (r = .80, p = .001); Interest (r = .61 p = .005) and with Ecological validity (r = .49, p = .03). Ecological validity was significantly positively associated with Spatial presence (r = .46, p = .048) and Interest (r = .65, p = .003). Both Engagement and Ecological validity were significantly positively associated with how strongly participants empathised with the characters (Engagement: r = .71, p = .001; Ecological validity: r = .67, p = .002), but not with Identification with the characters, although Empathy and Identification were themselves significantly positively linked (r = .63, p = .004).

8.4.4.2 TTS AD: Fear

For Fear clips with TTS AD, Fear was only significantly (negatively) associated with one presence scale: Negative effects (r = -.59, p = .009). Engagement was significantly positively associated only with Confusion (r = .62, p = .006). Ecological validity was significantly positively associated with Spatial presence (r = .70, p = .001) and Fear (r = .48, p = .043). Empathy and Identification were significantly positively linked (r = .57, p = .013) but had no significant correlations with presence.

Is there a significant correlation		**No AD (fear)**	**TTS AD (fear)**	**HV AD (fear)**	**No AD (sad)**	**TTS AD (sad)**	**HV AD (sad)**
… between affect and Spatial presence?	**Identification**	No	No	No	No	No	Yes*
	Empathy	No	No	Yes*	No	Yes*	No
… between affect and Engagement?	**Identification**	No	No	No	Yes**	No	Yes**
	Empathy	Yes**	No	No	Yes**	Yes*	Yes**
… between affect and Ecological validity?	**Identification**	No	No	No	Yes**	No	Yes**
	Empathy	Yes*	No	Yes*	Yes**	Yes*	Yes*
… between affect and Negative effects ?	**Identification**	No	No	No	No	No	Yes*
	Empathy	No	No	Yes**	No	No	Yes**

*. Correlation is significant at the 0.05 level (2-tailed).
**. Correlation is significant at the 0.01 level (2-tailed).

Table 8.4 Summary of correlations between Identification, Empathy and presence scales (significant correlations shaded in grey)

8.4.4.3 HV AD: Fear

For Fear clips with HV AD, there was no correlation between the target emotion and presence. However, Spatial presence was significantly positively linked to Interest (r = .58, p = .013) and Identification (r = .56, p = .014). Identification was also linked to Ecological validity (r = .483, p = .036). There was a positive association between Empathy and Negative effects (r = .65, p = .003). All three positive Presence scales were themselves significantly linked: Engagement and

Ecological validity (r = .55, p = .015); Engagement and Spatial presence (r = .47, p = .043); Ecological validity and Spatial presence (r = .80, p < .001).

Engagement was significantly associated with how strongly participants identified with the characters (r = .55, p = .014). Ecological validity was significantly positively linked to Empathy (r = .49, p = .035). Empathy and Identification were themselves significantly positively linked (r = .63, p = .003).

8.4.4.4 No AD: Sadness

As with the Fear clips in the No AD condition, bivariate correlations showed that, for the Sadness clips, the target emotion (Sadness) was significantly positively associated with Engagement (r = .60, p = .007). Sadness was also significantly positively associated with Ecological validity (r = .70, p = .001). Engagement and Ecological validity were themselves positively and significantly linked (r = .80, p < .001), and both were significantly associated with Interest (Engagement: r = .58, p = .009; Ecological validity: r = .48, p = .04) as well as Spatial presence (Engagement: r = .51, p = .028; Ecological validity: r = .56, p = .013).

All three positive presence scales were significantly positively associated with Empathy (Spatial presence r = .46, p = .047; Engagement r = .61, p = .005; Ecological validity r = .64, p = .003). Identification with characters was significantly linked to Engagement (r = .65, p = .003) and Ecological validity (r = .64, p = .003). Empathy and Identification were themselves significantly positively linked (r = .77, p < .001).

8.4.4.5 TTS AD: Sadness

For Sadness clips with TTS AD, there was no correlation between the target emotion and presence. Interest was significantly positively associated with Spatial presence (r = .46, p = .045), and all three positive dimensions of presence were linked: Spatial presence and Engagement (r = .88, p < .001); Spatial presence and Ecological validity (r = .81, p < .001); Ecological validity and Engagement (r = .76, p < .001).

Engagement and Spatial presence were significantly positively linked to Empathy (Engagement: r = .70, p = .001; Spatial presence: r = .60, p = .011) and to Identification (Engagement: r = .64, p = .003; Spatial presence: r = .50, p = .030). Empathy and Identification were significantly positively linked (r = .70, p = .001).

8.4.4.6 HV AD: Sadness

In the HV AD condition, the target emotion (Sadness) was significantly positively associated with Engagement (r = .60, p = .007). Again, all three positive dimensions of presence were significantly linked: Spatial presence and Engagement (r = .72, p = .001); Spatial presence and Ecological validity (r = .83, p < .001); Ecological validity and Engagement (r = .74, p < .001).

Engagement and Ecological validity were significantly positively associated with Empathy (Engagement r = .63, p = .004; Ecological validity r = .53, p = .020) and Identification (Engagement r = .58, p = .010; Ecological validity r = .63, p = .004). Identification was also significantly linked to Spatial presence (r = .51, p

= .027). Empathy and Identification were negatively linked to Negative effects (Empathy: r = -.54, p = .018; Identification: r = -.57, p = .011) and were themselves significantly positively linked (r = .89, p < .001).

8.4.4.7 Comparison of Fear and Sadness

Paired samples t-tests (Table 8.5) for the No AD condition showed ratings of Interest, Sadness, Empathy and Identification were significantly higher for the Sadness clips than the Fear clips. That pattern was reversed for Fear and Confusion. There was no difference between the 2 categories (Fear/Sadness) for Arousal or Tension. This was replicated whether comparing raw scores or the increase from baseline. There was also no difference between categories for Spatial presence. However, Engagement and Ecological validity were significantly higher for the Sadness clips. Negative effects were significantly lower. As the Sadness clips were generally considered more affecting and less confusing (possible reasons for this are discussed below) it is perhaps not surprising that the addition of AD did little to boost levels of presence or emotion. However, equally, AD did not lower emotion or presence levels.

Within groups, in the No AD condition, Fear clips differed on only one measure: Spatial presence: *The Shining* provoked significantly more than *Lambs* (mean dif = 1.65, p = .039). For the Sadness clips, *Four Weddings* was significantly more engaging than *Bambi* (mean dif = 1.73, p = .015).

	No AD (fear)	No AD (sadness)	t	df	Sig.
Arousal	4.31 (3.25)	5.36 (2.01)	-1.32	18	n.s.
Interest	4.63 (2.87)	6.32 (1.89)	-2.43	18	.03
Confusion	5.79 (2.92)	3.79 (2.95)	2.77	18	.01
Fear	3.26 (2.89)	1.26 (1.76)	2.94	18	.01
Tension	5.21 (2.62)	4.79 (1.51)	.72	18	n.s.
Sadness	1.53 (2.09)	6.11 (2.42)	-5.89	18	< .001
Empathy	3.21 (1.99)	5.74 (1.76)	-4.52	18	< .001
Identification	2.73 (1.82)	5.47 (1.87)	-6.25	18	< .001
Spatial presence	3.00 (1.16)	2.83 (1.36)	.435	18	n.s.
Engagement	2.89 (1.26)	3.87 (1.18)	-2.49	18	.02
Ecological validity	2.32 (1.27)	3.77 (1.13)	-3.91	18	.001
Negative effects	1.74 (1.15)	1.21 (.713)	2.38	18	.03

Table 8.5 Planned comparisons of raw emotion and presence mean scores (SD) for Fear and Sadness clips in the No AD condition

8.4.4.8 Gender

Across the three conditions, Chi square tests showed there was no effect of gender on levels the target emotion generated in either category: Fear (No AD: p = .61; TTS AD: p = .72; HV AD p = .43); Sadness (No AD: p = .60; TTS AD: p = .20; HV AD: p = .52).

8.4.4.9 Effects of sight characteristics on emotion ratings

For Fear clips, a repeated measures ANOVA comparing visual acuity (no useable vision N = 10; some useable vision N = 9) showed a significant interaction only for Confusion: $F(2,32) = 4.77$, $p = .015$. Those with some useable vision found No AD significantly more confusing than TTS AD or HV AD. Those with no useable vision found TTS AD and No AD equally confusing.

There were no significant main effects or interactions for those with visual experience compared with participants with no visual experience. However these results should be treated with caution as there were only 5 participants in the latter category.

For Sadness clips there were no significant main effects or interactions either between those with no/some useable vision or between those with and without visual experience.

8.5 Results: Physiological data

For ease of manipulation, data for each physiological measure (HR, HRV and EDA) were averaged and binned in 10-second blocks. The resulting data was manually inspected. Single dropouts were smoothed by averaging the data points from the 10s time periods either side. Where the data included more than one drop out in a row, that participant was excluded from the analysis. In all 7 participants were excluded and 12 retained. Scores for each measure were calculated, clip by clip, using a global baseline (i.e. by subtracting the mean of the original 30s baseline period from the raw scores for each measure). Mean scores for each condition (No AD/TTS AD/ HV AD) were then compared on a within-participants basis. As, for example, HRV tends to increase with excited feelings but decrease in response to sad stimuli (Ridgeway & Waters, 1987) the two emotion categories were analysed separately.

For Fear clips, a repeated measures ANOVA showed that, with Greenhouse-Geisser corrections, condition (No AD/TTS AD/ HV AD) was significant only for HRV: $F(2, 21.57) = 4.84$, $p = .046$. Planned paired comparisons showed that HRV was significantly lower ($p = .042$) in the No AD condition ($m = -1.07$) than for TTS AD ($m = .785$). The difference with HV AD ($m = -.083$) was not significant: $p = .130$. Given the small number of participants the 3 conditions were also compared using a Friedman's test. This was also significant ($p = .032$) for HRV. Sample traces for Fear clips for one participant (difference from baseline) are shown in Figs. 8.10 – 8.12. For Sadness clips, there were no significant main effects of condition on any of the physiological measures.

8.5.1 Physiological correlations: Fear

A summary of bivariate correlations is given in Table 8.6 (see Appendix C6 for more detail). In the No AD condition, EDA was significantly negatively related to HR ($r = -.68$, $p = .015$). HR was also significantly negatively associated with HRV ($r = -.69$, $p = .002$). However, none of the physiological measures (HR/HRV/EDA) were significantly associated with the target emotion (Fear) or with any other emotion on the EES. There were also no significant associations between physiological measures and any of the presence scales.

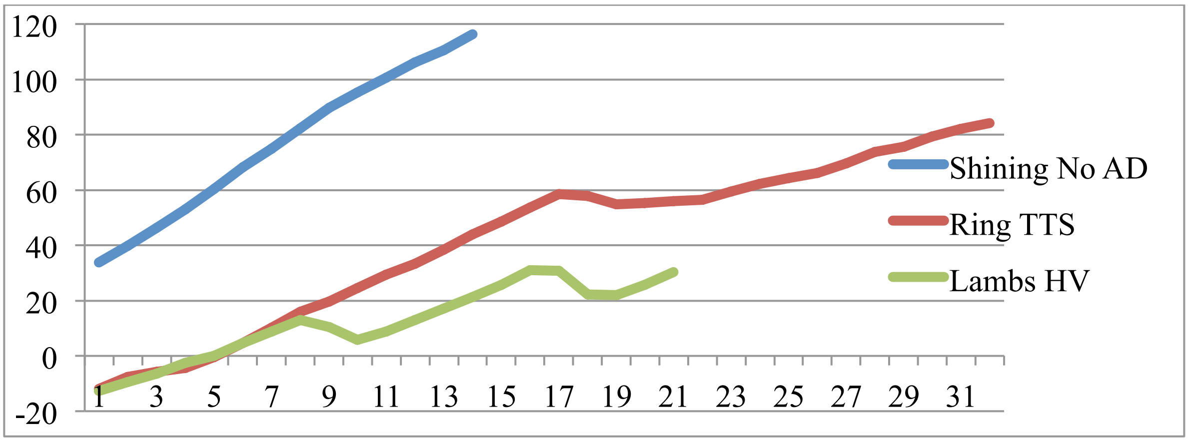


Fig. 8.10 Sample traces of EDA (difference from baseline) for participant E03

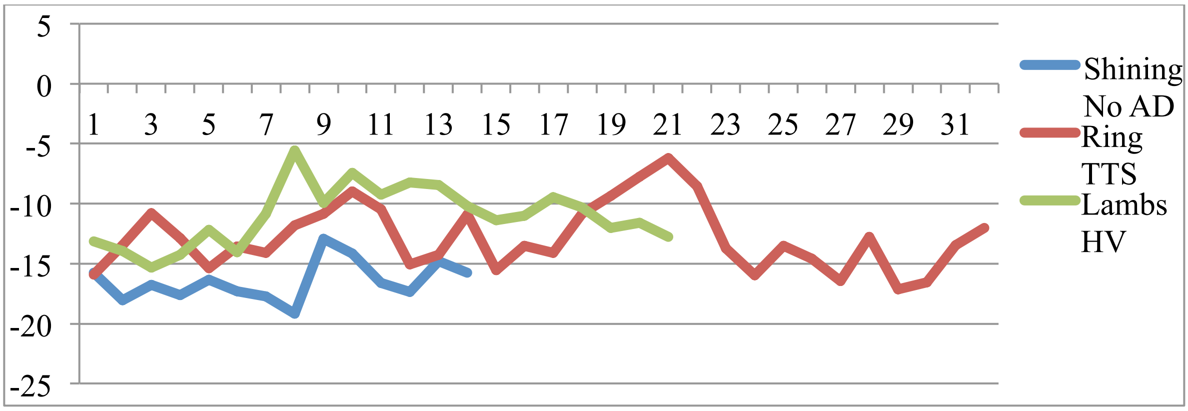


Fig. 8.11 Sample traces of HR (difference from baseline) for participant E03

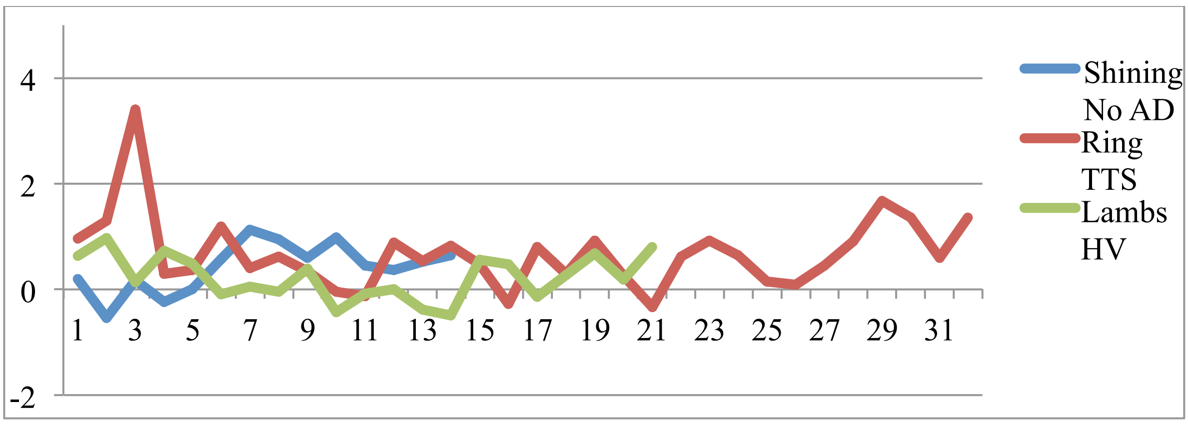


Fig. 8.12 Sample traces of HRV (difference from baseline) for participant E03

However, EDA was significantly positively associated with Identification with the characters (r = .66, p = .018).

Are there significant correlations between…	**No AD (fear)**	**TTS AD (fear)**	**HV AD (fear)**	**No AD (sad)**	**TTS AD (sad)**	**HV AD (sad)**
… all 3 physiological measures?	No	No	No	No	No	No
…physiological measures and target emotion?	No	No	No	No	No	No
…physiological measures and Confusion?	No	No	No	No	Yes* (HRV) Yes* (HR)	No
… physiological measures and Identification?	Yes* (EDA)	No	No	No	No	No
…physiological measures and Empathy?	No	No	No	No	No	No
… physiological measures and Spatial presence?	No	No	No	No	No	No
… physiological measures and Engagement?	No	No	No	No	No	No
… physiological measures and Ecological validity?	No	Yes* (HR)	No	No	No	No
… physiological measures and Negative effects?	No	Yes* (HR)	No	No	No	No

*. Correlation is significant at the 0.05 level (2-tailed).
**. Correlation is significant at the 0.01 level (2-tailed).

Table 8.6 Summary of correlations between physiological measures, emotion and presence.

In the TTS AD condition, HR was significantly negatively linked to Arousal ($r = -.51$, $p = .03$) and EDA was positively linked to the non-target emotion, Sadness ($r = .64$, $p = .018$). HR was positively associated with Ecological validity ($r = .48$, $p = .05$) and negatively linked with Negative effects ($r = -.54$, $p = .021$). There were no significant correlations between any of the physiological measures and Identification or Empathy.

In the HV AD condition, there were no significant correlations.

8.5.2 Physiological correlations: Sadness

In the No AD condition, HR and EDA showed a significant positive association ($r = .49$, $p = .048$). However there were no significant correlations between any of the physiological measures and emotions on the EES, any of the presence scales, Identification nor Empathy.

In the TTS AD condition, there was a positive correlation between HR and HRV ($r = .73$, $p = .003$). There were no significant correlations between physiological measures and the target emotion, but significant positive associations between HR and Confusion ($r = .65$, $p = .011$) and HRV and Confusion ($r = .58$, $p = .012$). There were no significant correlations between the physiological measures and any of the presence subscales, or with Identification or Empathy.

In the HV AD condition, there was a negative correlation between HR and HRV ($r = -.57$, $p = .018$). There was a significant negative correlation between EDA and the non-target emotion, Fear: $r = -.55$, $p = .027$. There were no significant correlations between any of the physiological measures and presence subscales, nor Identification or Empathy.

8.6 Discussion

This is the first study, as far as the author is aware, to compare links between emotion and presence in participants with a major sensory disability. Its aim was to test whether particular dimensions of presence are associated with greater experienced emotion and, in particular, whether AD has an impact on emotion elicitation, empathy and presence.

Specifically the study tested the hypothesis that, through the mechanism of auditory compensation, B&PS people could successfully access emotions from the soundtrack (music, SFX and dialogue) and that AD, by reducing access to that soundtrack, would lead to a reduction in experienced emotion. As well as comparing the experience of "watching" emotive movie clips with and without AD, the study also compared AD delivered by a human voice, and the same script delivered by an electronic voice devoid of affective paralinguistic content. Results revealed major differences in response to the two emotion categories, Fear and Sadness. However, in neither was there evidence that the target emotion was elicited more strongly in the No AD condition compared with adding AD.

8.6.1 Target emotion

The first question was whether clips known to elicit particular emotions in sighted people would have the same effect on people with a visual impairment. For Sadness clips with No AD, the target emotion (Sadness) was successfully elicited significantly more strongly than other emotions. This was also the case when those same clips were watched with TTS and HV AD. However, Fear clips without AD failed to elicit the expected response i.e. the target emotion (Fear) was experienced no more strongly than any other emotion, with the exception of Interest. Arguably, without access to visual information, Interest was particularly low. This might account for lack of Fear. Yet the same result was found for TTS AD which replaced missing visual with verbal information. Only with HV AD did the clip achieve its effect: Fear was elicited significantly more strongly than any other emotion except for the related emotion of Tension. Tension, like Fear, is a key constituent of horror films, and a strong correlation has previously been found between the two emotions (Dillon, 2006) so this was not surprising.

Fear was not only experienced more strongly than other emotions while watching with HV AD, it was experienced at a significantly higher level than when watching the clips without AD. Participants also found the HV AD clips significantly more arousing. There was no support then for H1, that No AD would elicit higher levels of the target emotion than AD. However, there was only partial support for H2. Levels of the target emotion were only higher with AD in one emotion category (Fear) and then only with HV AD. This meant that H3 (TTS AD < HV AD) was also only partially supported.

8.6.2 Media content: Comparing emotion categories

Differences were clearly apparent between the two emotion categories, Sadness and Fear. In the No AD condition, participants found the Sadness clips less confusing and more emotive than the Fear clips (with the exception of Fear itself). Participants also showed greater empathy with characters in the Sadness clips and identified with them more strongly. This underlines the importance of media content in emotion elicitation. For those with a visual impairment, content can be thought of not only in terms of subject matter but also the relative balance between auditory and visual information. It was particularly noticeable that the Fear clips were difficult to follow with sound alone. *The Shining*, for example, is underscored throughout by non-diegetic music and has only a couple of short interjections of dialogue. A little boy, Danny, playing alone in an empty hotel corridor is surprised when a tennis ball rolls towards him, apparently from nowhere. He calls out "Mom" and approaches the half-open door to a hotel bedroom: "Mom…Mom are you in there?" The final 50s of the clip is solely music, while the visuals show the empty bedroom, a hand pushing open the door of the en suite bathroom, a figure in the bath partially glimpsed through the shower curtain, and the face of Danny's Dad, Jack, who stands in the bathroom doorway watching with terror as the curtain is slowly drawn back. Without the visuals, the dialogue is potentially misleading. Jack does not speak. Neither does the mysterious figure in the bath. There is no audible indication that they are there. Only Danny has a voice. It is possible, then, to assume the clip concerns a little boy looking for his mother. The music is scary and includes an amplified heartbeat that could easily be associated with the anxious child. However, Danny is soon left out of shot, as the camera takes us into the empty bedroom. The heartbeat helps build suspense for the sighted audience as they glimpse the shadowy figure in the bath, whose presence cannot be detected without sight.

By contrast, the Sadness clips were largely self-explanatory, with a greater proportion of coherent dialogue. However, there were still elements that caused confusion. For example, there is no dialogue for the first 65s of *Weddings*. The first location is a suburban housing estate, where a funeral car draws up to collect Gareth's bereaved parents and his partner Matthew. The scene then moves to a Church where mourners are waiting in the pews. This is only apparent from the soundtrack once the Vicar begins to speak. Even then this might be just a normal church service, until, 80s from the start of the clip, the Vicar invites Matthew to address the congregation. Interestingly, although HV AD significantly increased understanding, it gave no additional benefits of emotion elicitation.

8.6.3 Understanding and emotion elicitation

For the Fear clips both HV AD and TTS AD made the clip significantly less confusing but only HV AD significantly heightened the sense of fear. This suggests that the paralinguistic emotional content played a role with fear conveyed through the delivery rather than the words of the describer. In no condition for either emotion category was there a correlation between confusion (or lack of it) and the target emotion. Evidently, then, understanding every detail of what is going on is not essential to experience the emotional content. It is particularly interesting to note that the addition of visual information, such as Bambi's eyes filling with tears when the Stag tells the little fawn that his mother is never coming back, made no significant impact even on LB/PS participants who might be expected to gain most from access to bimodal reinforcement. This corresponds to Putzar et al.'s (2007) finding that, when adding a congruent visual stimulus to an auditory one, there were no redundancy gains for individuals who had been born with congenital cataracts and had had their sight restored.

8.6.4 Social presence

If understanding is not the key to unlocking the difference between emotion categories, it may be helpful to look to the ratings of presence, Identification and Empathy. In the No AD condition, Sadness clips showed significant correlations between all three positive dimensions: Spatial presence, Engagement and Ecological validity. The same was true in both AD conditions, and when HV AD was added to the Fear clips. For all these stimuli the target emotion was successfully elicited, suggesting a clip might best achieve its emotional impact when it is cohesive i.e. triggering all dimensions of presence (as measured by the ITC-SOPI). Yet there was no direct correlation between the target emotion and Spatial presence in any condition for either emotion category. This is perhaps understandable for the Sadness category. The sadness you feel listening to Matthew's speech at the funeral, for example, is unlikely to be enhanced by details of the church architecture, but it is surprising that feeling more present in the gloomy basement during the chase sequence in *Lambs* does not coincide with a heightened sense of Fear.

It suggests that emotion might be more strongly influenced by Empathy and Identification with the characters. Certainly the clips that successfully elicited their target emotion were those that showed most correlations between Empathy, Identification and the various dimensions of presence. In the HV condition there was also a significant correlation with Negative effects. This was positive for the

Fear clips suggesting that sharing a nasty experience (Fear) was itself unpleasant. By contrast, for the Sadness clips the correlation was in the opposite direction, as though sharing the experience with the bereaved character diminished Negative effects.

As a correlational link is not directional, the question arises whether a greater sense of presence stems from greater emotion elicitation or the other way around. Arguably the more the participants identified with the characters' situation and empathised with them, the more present they felt themselves to be, with the scenes seeming natural and the clip overall being more engaging. This suggests that where participants were able to draw on their own experience, they could feel the target emotion more strongly. This is evidently more likely for a clip dealing with bereavement (*Bambi*, *Truly*, *Weddings*) than one confronting a mass murderer (*Lambs*) or threats from the supernatural (*The Ring*, *The Shining*).

It is possible that the significantly higher ratings of presence reported for the Fear clips presented with HV AD arose from an increase in Social presence. Just as Mandryk et al. (2005) showed that playing a computer game against a friend, as opposed to playing against the computer, induced a greater sense of presence, so experiencing a scene in the company of a human voice rather than an electronic one, may have had a similar effect. By giving the impression of a "shared" experience, HV AD may enhance psychological "transportation", in line with findings for sighted people (Shedlosky-Shoemaker et al., 2011). It may also seem more real, reflecting occasions on which, when watching TV with no AD, sighted family or friends improvise a descriptive commentary. Future research might explore whether Social presence might be greater in live AD settings (where the describer is actually present) compared with recorded HV AD for screen-based media.

8.6.5 Physiological measures

Physiological measures showed no consistent pattern either between emotion categories or within. Heart Rate Variability (HRV) was the only measure to show a significant response. For Fear clips alone, it was significantly higher for TTS AD than No AD or HV AD. It might be assumed, then, that the Fear clips were scarier with TTS AD, yet this conflicts with the self-report measures. There are two explanations. The first is one of social desirability. Most participants knew that the researcher was an audio describer and realised that the HV AD used the researcher's voice. It may be that they claimed, out of politeness, to be more afraid with HV AD. There are several reasons for thinking this was not the case. If it were true, it is surprising that they did not also claim to find the Sadness clips sadder with HV AD. Moreover, in comments after the experiment was completed, several participants talked about their experience of electronic versus human speech. One congenitally blind man did not recognise the voice of the describer. He said:

> "I like the way the (HV) voice lifts at the end of a phrase – it's softer. You want to ignore the describer. An electronic voice spoils it for me. I prefer a natural voice 100%. It's hard to explain how much difference it makes. I heard that (TTS AD) and I thought "Oh God, no". It's not as clear which means you have to concentrate on the describer, which you

> don't want to do. And it sounds colder. We like human beings." (E10)

His opinion was not universal. One partially sighted man said: "With the electronic AD I was not as detached as I thought I might be. It was OK" (E08). However, the fact that TTS AD was harder to understand was echoed by several participants. E04 said:

> "When you're listening to TTS it's harder to understand, so you have to concentrate more. Especially when there's a lot going on, it's harder to pick up the thread than with human AD. TTS is pulling you away from the scene."

E09 expressed it this way:

> "TTS makes a difference to me with comprehension. Also it's got a tone so you can't put your interpretation on it. It's not neutral. I definitely felt disconnected. It's a human voice gone wrong. It's fine for factual books but it ruins fiction."

The alternative explanation for TTS AD eliciting higher HRV for the Fear clips then lies in the fact that although HRV is known to correlate with fear, arousal and vigilance, it also increases with excessive cognitive load and mental frustration (Kim et al., 2004). Possibly HRV was higher because TTS took too much cognitive effort to decode. In particular, B participants found TTS AD significantly more confusing than PS participants. The former group included EB and CB, but visual experience was not a contributory factor to ease of understanding. It may be that people with some useable vision needed to rely less on the AD.

For the Sadness clips, TTS AD showed a significant correlation between HRV and Confusion, although it should be noted there was no significant drop in presence. HRV also showed a significant correlation with HR in this condition. Wilson & Sasse (2000) found that HR increased significantly when the visual stream of a video-conferencing call was reduced to 5 frames per second (rather than the usual 25 frames). Their research also showed increased HR in conditions of poor audio quality. Again this suggests that increased HR (and by extension HRV) may arise from frustration or excessive cognitive load.

Interestingly, given the evidence that, for S, verbal information carries a higher cognitive load than processing perceptual information (Klatzky et al., 2006) none of the B&PS participants reported being overwhelmed by the AD. This supports those studies that suggest B&PS have increased WM capacity for verbal information (Raz et al., 2007; Rokem & Ahissar, 2009). Although Rokem and Ahissar specifically attributed this to a higher threshold in the ability to discriminate speech from noise, it would further suggest that intelligibility of speech is a key factor that would be worthy of future research.

Overall, the addition of AD elicited an emotional experience that was, at worst, equal to watching with no AD and, at best, significantly more emotive. This suggests that the addition of verbally-presented visual information does not detract from emotional engagement and can positively increase it. Furthermore, greater emotional engagement is linked to higher ratings of presence. In terms of applying these findings to the practice of AD, the results offer conflicting evidence but it seems that what counts, as much as *what* you say, is *how* you say it.

8.7 Limitations

The lack of consistent pattern across physiological measures demonstrates the difficulty of assessing emotional response. One blind participant spontaneously commented that there are good and bad types of tension: good when induced by the narrative, bad when it arises from the frustration of not knowing what is happening. Another potential confound stems from individual differences depending on an individual's ability to cope with stress. Those with a quicker PNS response (i.e. the ability to bring the HR quickly back to normal), who are better at regulating their emotions, would show higher HRV compared with less adaptable peers (Appelhans & Luecken, 2006). Conversely, a lower HRV has been associated with an inability to generate appropriate emotional responses (Sgoifo et al., 2003). Such differences led to the adoption of a within-participants analysis, which limited the extent of the study. It was not possible to examine differences in response to the same clip presented in different conditions. Clips were therefore grouped in emotion categories. Within categories, however, clips still varied in terms of content. Furthermore, content varied even within a single clip. Most clips conveyed more than one type of emotion or were easier to follow without AD at some points more than others. The small sample size and the low number of participants with no visual experience also meant comparative results between CB/EB and LB/PS should be interpreted with caution.

Physiological data offers the possibility of comparing responses to specific incidents in the content. Mandryk et al. (2007), for example, compare reactions to a goal being scored. However, it was difficult with the clips in the current experiment to pinpoint exact moments within the narrative flow. For example, Fear and Tension build gradually in *The Shining*, without there necessarily being a single "scary" moment. This problem was exacerbated by the fact that, for ease of manipulation, physiological data was binned in 10s blocks. This may have smoothed out potential spikes and troughs. A smaller block-size and greater accuracy in video time codes would be useful in future experiments.

8.8 Summary of Study 6

Despite the caveats above, this study showed that the addition of verbal information did not lead to a reduction in presence or levels of elicited emotion, despite the AD partially masking the soundtrack. This is in line with Study 4 in which SFX were shown to make little difference to B&PS people's sense of presence. It suggests that SFX and emotive music may be redundant where sufficient information can be accessed from the dialogue (including paralinguistic information such as the characters' tone of voice). Equally, visual information carried in the AD that reinforced an audible emotional response also proved redundant. In the Sadness clips, the visual details did nothing to enhance the emotional impact of the scene, whether or not participants had residual vision or visual experience. In the Fear clips, where the dialogue did not provide sufficient information, SFX and music allowed participants to recognise the emotion the clip was eliciting, but did not allow them to experience it directly. Verbal information lessened confusion. However, the target emotion could only be fully experienced with HV AD. The link between physiological measures and levels of Confusion suggest that the complexity of decoding an electronic voice may be one reason why TTS AD did not achieve the same effect. However, the higher levels of presence associated with HV AD has also been explained in terms of Social presence: "sharing" the experience with a human voice leads to greater empathy, a stronger emotional experience and consequently greater Engagement and a feeling that the mediated environment is real.

8.9 Implications for Audio Description

In their comparison of AD guidelines across countries, Rai et al. (2010, p. 71) suggest that the greatest challenge is choosing "what not to describe". This is discussed in more detail in Appendix D in the context of the thesis as a whole. However, this Study provides a number of general points that providers of AD might consider. Most guidelines agree that the main aim of AD is to help users follow the plot. In this Study Confusion was reduced by AD. However it was also clear that Confusion was not directly related to enhanced presence or emotion elicitation. For example, Fear clips were significantly less confusing both with HV and TTS AD, but only the human voice elicited significantly higher levels of the target emotion. This suggests that AD approaches need to be considered in relation to the aim of the source content. Content designed for entertainment may require a different style of AD, or at least AD delivery, from content with an educational role.

Guidelines offer conflicting advice in terms of delivery style, i.e. whether the AD should be delivered objectively or in sympathy with the scene. Again, this may need to vary to reflect the source material. In this Study HV delivery reflecting the emotional content enhanced not only semantic but also emotional understanding. In the light of this, moves to increase AD output by use of TTS, as currently being implemented for example by the company Swiss TXT (Linder & Martinez, 2011), should be carefully assessed. It may be that TTS is suitable for factual programme genres but less so for drama.

Igareda and Maiche (2009) point out that common to all AD guidelines is a lack of scientific research to underpin current practice. Studies such as this one highlight how sighted describers may benefit from a greater understanding of the processing style of their audience when choosing what visual information to include. Bimodal reinforcement of auditory cues with visual information (verbally expressed) was shown not to enhance presence or emotion elicitation. Given the attentional and cognitive demands of AD, it may be better, then, at such times to let the soundtrack to do its job. However, when the soundtrack provides insufficient or incongruent information then the visuals can be described without concern that AD will negatively affect presence. Evidence from this study suggests this is best achieved with AD delivered by a human voice.

More research is clearly needed to clarify whether the germane load of AD content can be determined in terms of quantity and style. The intrinsic load may be lightened by more minimal AD or by a simpler vocabulary or word structure. However this is likely to depend on the educational background and sight characteristics of the AD user and their interest in the visual world. In particular, given the comments in Chapter 3, minimal AD may result in the user feeling patronised or excluded. It may be simpler as a first step to explore the difference in lightening extrinsic load by lowering the level of competing background sound for the duration of the AD and ensuring the AD itself is more intelligible.

8.10 Presence as a measure of AD quality

The small amount of published research into AD reception (e.g. Peli & Fine, 1996; Schmeidler & Kirchner, 2001; Szarkowska & Jankowska, 2011) has tended to rely on AD-specific questionnaires. As participants in Study 1 made clear, however, the best AD goes unnoticed. This "perceptual illusion of non-mediation" (Lombard & Ditton, 1997) suggests that, in addition to using presence as a measure of user experience, it may also be used to measure the efficacy of AD itself. In the current study, for example, clear differences were revealed between TTS and HV AD. Presence measures suggested that the AD was less noticeable (i.e. participants felt more engaged and that the content was more natural) when delivered in a human voice. Similar findings have recently been reported for a study involving visually impaired children (Zabrocka-Suwka, 2013).

Clearly there are limitations. Existing presence measures are designed for virtual reality environments where users have a degree of control over their immersive experience. At present, only a handful of interactive, virtual environments have been designed with blind people in mind (Rai et al., 2010). Self-report measures include statements such as “I felt able to change the course of events”. This is unlikely to apply to the type of media for which AD is currently provided, such as TV and film that rate low on a scale of interactivity. As mentioned in Chapter 6, current measures may also need modification to make them appropriate for visually impaired participants. Statements such as “I had eyestrain” may not be appropriate for some blind users. The design and validation of more universal measures would be a welcome addition to presence research and to research in AD.

As with any post-hoc, self-report questionnaires, presence measures are subject to demands of social desirability: respondents report what they think the researcher wants to hear. This is all the more likely when participants with a visual impairment have to provide an oral response. In Study 4 this was avoided by using an online survey, which blind people could access by a screen-reader. This had the downside that sound quality of the stimuli could not be monitored. Study 6 used the haptic slider to avoid participants having to make an explicit response. However, participants may still have been guarded as the researcher was obliged to mark the score sheet. Post-hoc measures also rely on participants’ memory for the experience and do not allow for changes in presence across the duration of a dynamic scene. This was true even of the relatively short clips in this study. In AD settings of long duration, such as a film or a play, this is of particular concern.

Study 6 aimed to overcome this by using physiological measures. However data dropout, high individual variability, and difficulties in interpreting results made it less helpful than was hoped. A system of continuous monitoring would be a useful development, although this has practical difficulties and generally facilitates participants reporting on only one dimension of presence at a time. Nonetheless, improving ease of use of physiological measures would be a helpful step forward, particularly for emotive content.

Despite these limitations, presence measures have the advantage that they are already used widely across broadcast, advertising and virtual-reality industries and in academic research. Minor modifications (such as substituting the word “scene” for the word “display”) are unlikely to substantially alter their validity. They are reliable and multi-dimensional, and, as the above study illustrates, assess the entire experience of the AD user, rather than their experience of the descriptive element alone. Designed for audiences in general, presence measures have the added advantage that the experience of AD users with a variety of sight conditions can be compared directly, not only with each other, but also with the experience of the sighted audience. This enables the efficacy of the AD to be gauged in line with its stated aim: to provide access to mediated environments.

Chapter 9 General Discussion

9.1 Introduction

This thesis has presented a series of studies showing the impact of visual impairment on perception and experience both in the real world and in mediated environments. While previous research has often focused on people with congenital blindness, the studies presented here used participants with a range of sight characteristics to reflect the demographic of the blind population. The intention was to extend current models of presence to an audience with a major sensory disability. The studies addressed a number of research questions: What role does visual experience play in shaping the imagination? How does the brain adapt when vision is absent or limited? In particular, what mechanisms might compensate for a missing or intermittent stream of visual information in the real world and, by extension, the virtual world? This Chapter will review the theoretical models of the impact of visual impairment on cognitive processing before moving on to evaluate the findings of the studies with reference to those models. It will also consider the implications for presence and suggest avenues for future research.

9.1.1 Theoretical background

Chapter 1 outlined current models of perception. It emphasised the continuous interaction between direct sensory information and stored knowledge in order to understand the world, and the constant shift in balance between external cues and internal, task-driven goals. In particular it focused on Gibson's ideas of affordance and the role of perception in recognising what the environments, objects and people around us have to offer. While Gibson largely concentrated on visual perception, the Chapter reflected the growing awareness of the importance of multisensory information pick-up. The brain must integrate data from different senses to produce a coherent understanding of the world.

9.1.2 Visual dominance

Sight is generally acknowledged to provide the overarching reference frame within which non-visual sensory data is organised (e.g. Röder, Focker et al., 2008; Spence & Gallace, 2011b). Sight has a tendency to dominate bimodal, and presumably multimodal, interactions outweighing information from other senses. This is demonstrated by illusions such as the ventriloquist effect and the rubber hand illusion. The dominance of sight is important in a mediated environment. If a scene is visually engaging, it helps us disregard information from non-visual senses (an uncomfortable chair, coughing from a fellow audience member, the stuffy air in the theatre) that would otherwise disrupt our sense of presence by bringing our attention back to the real world. Yet recent studies have shown that a poor visual display matters less than poor sound quality when it comes to accepting virtually presented objects as real (e.g. Skalski & Whitbred, 2010; Bonneel et al., 2009).

9.1.3 Sensory compensation

A heightened response to sound, together with other non-visual senses, is thought to be the main mechanism by which people who are blind or partially sighted compensate for lack of sight. This results from crossmodal plasticity, by which connections in the brain are reorganised, either in the absence of visual input as the brain develops or later as the result of experience, through greater reliance on non-visual modalities. Röder and Rösler (2004) have proposed that if non-visual senses compensate for lack of vision, CB should out-perform sighted people in non-visual tasks. Cattaneo and Vecchi (2011) extend this logic to differences within the blind population. They argue that if visual experience is necessary for normal cognitive processing, then CB will show impaired performance in all modalities compared to their EB/LB/PS peers. However, if senses develop independently, CB/EB will show impaired performance only in tasks requiring visual memory. Given that current evidence for sensory compensation is inconsistent, this thesis considered three other possibilities:

i. If only CB and EB benefit from cortical reorganisation, they will outperform LB/PS in tasks involving non-visual modalities.
ii. If compensatory plasticity continues throughout life, driven by experience, difference in performance between CB, EB, LB & PS will vary in accordance with extent and duration of sight loss.
iii. If experience compensates for lack of early brain plasticity the performance between CB, EB, LB & PS may be similar but result from different strategies or processing systems.

More anecdotal evidence, advanced by Oliver Sacks (2003), was also considered. Namely that the categorising of B&PS according to their aetiology of sight loss overlooked individual differences in interests, experience and cognitive styles, in particular with regard to imagery.

9.1.4 Crossmodal integration, mental imagery and linguistic compensation

With little or no access to vision as an organising framework, it has been suggested that visual impairment results in individuals being less good at multisensory integration (e.g. Hötting et al., 2004). This seems to be based on results of tasks that are inherently affected by visual experience (e.g. the crossed-hands paradigm). Yet if blind people are to integrate modalities effectively, what takes the place of sight? Given neuroimaging evidence discussed in Chapter 1, that visual cortex is recruited by blind (but not sighted) people to process speech, it was proposed in the introduction that language might provide an alternative. Language is considered by some to be purely abstract or amodal (e.g. Fauconnier, 1999) and by others to be rooted in experience and therefore processed by modality-specific cortices (e.g. Barsalou et al., 2008). With reference to spatial imagery, Struiksma et al. (2009) suggest a combination of the two: that amodal language interacts with modal information to produce a supramodal image that is more than the sum of its parts. The studies presented here suggest that whether language is amodal/multimodal/supramodal may be affected by access to visual information. For those with a visual impairment, language is used at encoding thus making sensory experience explicit. This in turn affects recall, with explicit information (i.e. words) triggering as rich a simulated perceptual experience as implicit information (i.e. perceptual cues). The evidence for this is evaluated in the next section, which specifically considers Studies 2, 3, 4 and 5. The implications for presence are then discussed, with reference to Studies 1 and 6 in section 9.3.

9.2 Evaluation of Results

9.2.1 Perception, experience and multisensory integration

Study 2 focused on the impact of visual impairment on crossmodal associations, translating the Bouba-Kiki experiment from the auditory/visual modalities into the auditory/haptic modalities by presenting spiky and rounded object pairs that could be touched but not seen. Participants were asked to decide which was Bouba and which was Kiki. If Röder and Rösler (2004) were right, with non-visual senses compensating for lack of vision, CB should show a more robust Bouba-Kiki effect (matching rounded forms with Bouba and spiky forms with Kiki) as they should have heightened auditory sensitivity and greater haptic experience. Furthermore they should show a more robust effect than their LB/PS peers. The results showed the opposite. 84% of S consistently chose in the expected direction, compared with 56% of LB/PS. The admittedly small number of CB participants (6) chose at the level of chance i.e. only 50% of them showed the effect. If this is representative, it suggests that access to visual experience affects crossmodal connections even between non-visual modalities.

As well as questioning the relative contribution of the auditory and visual elements to the traditional Bouba-Kiki experiments (that present people with a picture of a rounded/spiky shape), Study 2 provides behavioural evidence consistent with neuroimaging data that shows, in sighted people, tactile object-recognition activates bimodal visuo-haptic neurons in the LOtv (Lacey, 2009). Similarly, once a bimodal audio-visual association has been established, the visual cortex can be triggered by sound alone (Zangenehpour & Zatorre, 2009). Study 2 suggests that such associations persist even in the wake of visual impairment. LB people were more likely to respond like S than CB. Access to visual memory may therefore be a constituent part of any mechanism that compensates for lack of direct visual information in people who have had sight. Evidence from the PS participants, whose results were similar to CB rather than, as might be expected, to LB suggests that if vision is consistently unreliable such bimodal (audio-visual or haptic-visual) connections do not develop.

In explicitly asking participants about their strategy, Study 2 showed that the auditory element of the experiment triggered semantic associations in some people (e.g. associating Kiki with a name) and more abstract auditory associations (associating Kiki with shrill or staccato sounds and therefore the more jagged shape) in others. It was particularly interesting that auditory associations were not consistent in B&PS, suggesting lack of exposure to environmental correspondence between sound and vision. It may also reflect individual differences in imagery, cortical brain plasticity and experience.

Study 3 compared the effect on imagery of verbal versus non-verbal auditory triggers. The study was in two parts. Part 1 explored clarity of auditory imagery. If the auditory compensation hypothesis was correct, B&PS might be expected to imagine sounds more clearly than S. Part 2 explored the multisensory nature of imagery prompted by words and sounds. Given the findings of Study 2, it seemed likely that imagery would reflect differences in visual experience between S and B&PS. It was also hoped that the study would shed some light on the unimodal/multisensory integration debate.

Study 3 was conducted remotely via an online survey site. Although this brought its own limitations, it enabled 100 S and 39 B&PS people to take part. This makes it a comparatively large study of people with a visual impairment. Of the 43 studies reviewed by Cattaneo & Vecchi (2011) only one had greater B&PS participant numbers (Afonso, Blum et al., 2010: N = 48). Auditory imagery was measured using a relatively new scale, the CAIS (Willander & Baraldi, 2010). This was designed to counter flaws in existing measures. However, either the CAIS proved to be insufficiently sensitive, or auditory imagery by its nature is, for many people, an all-or-nothing experience, as 50% of all participants, regardless of sight characteristics, reached the ceiling score. For the purposes of this thesis, the interesting finding was that the 50% was consistent across all groups i.e. S, CB, LB and PS. There was no evidence of task superiority for B&PS. This supports the results of Study 2 and is in line with other studies that have questioned claims of compensation in those with a visual impairment in *all* aspects of audition (e.g. Lewald, 2004).

The CAIS comprises 16 verbal descriptions of sound. For Part 2 of the study, these 16 items were also recreated as sound effects (SFX). The process of doing so revealed a certain degree of visual dominance on the part of the CAIS developers, with items reliant on visual as well as auditory imagery e.g. the sound of opening a can of *soft* drink requires visual information, such as the logo on the can or the colour of the liquid, to know what the can contains. Certain sounds were necessarily less specific and others more so than their verbal equivalents, which may have had an effect on the imagery they evoked. Imagery was measured using a 5-point Likert scale, with participants rating the strength of imagery across 7 modalities. The results were unexpected. Asked to identify the SFX, B&PS were no more accurate than S, again questioning auditory compensation. Furthermore, while S reported a heightened response to the SFX (i.e. generating images that were significantly stronger and made up of more modalities) than to the verbal descriptions of those sounds, B&PS people did not.

As in Study 2, then, Study 3 suggested that the effect of language on mental imagery was affected by vision, not only in whether or not specifically visual imagery was present, but also how strong and multisensory that image might be. The number of modalities prompted by language was reduced for S but not B, who showed a multisensory response in both conditions. Interestingly, an fMRI study by Hüfner et al. (2009) showed multisensory activations in S took place in an eyes-closed state but not when their eyes were open in darkness. Conversely, B adopted a residual “interoceptive” state of multisensory activations whether their eyes were open or closed. In Study 3, for word cues in particular, the dominant modality varied for B participants according to the particular item (e.g. associated smells were dominant for “the operation of a washing machine” and bodily movements for “the rustle of leaves underfoot”). If imagery reflects direct sensory experience, it suggests that, rather than auditory information specifically providing compensation for lack of vision, a combination of non-visual modalities may step in, their weighting dependent on the stimulus itself. PS participants, like S participants, consistently rated vision as the dominant mode of imagery. Nonetheless, PS resembled their B peers in showing no difference in the number of modalities that contributed to an image, whether it was prompted by words or sounds.

Study 3, however, contained a flaw in its design. Participants were asked to identify the SFX (by typing in a description) before moving on to assess the imagery those sounds evoked. It was possible that, in both conditions, participants were responding to verbal descriptions. Leaving aside the question as to why, then, S people's responses were significantly different between the two conditions, the possible confound was addressed in Study 4. Study 4 was also designed to extend the research questions from "static" imagery to a dynamic virtual environment. On the basis that B&PS might still show a heightened response to sound, the environment chosen was purely auditory: scenes from an audio drama heard with or without SFX. Compared with an AV medium, this also ensured a coherent source of information in both conditions, allowing responses to be gauged with no danger of B&PS missing crucial information from the visual stream.

Although ranking low on the list of immersive environments, audio drama still has the potential to induce a sense of "being there". It was decided, therefore, to compare the two conditions (with/without SFX) using a well-validated measure: the ITC Sense of Presence Inventory (ITC-SOPI, Lessiter et al., 2001). Again, the experiment was carried out online. Of the 21 B&PS only 3 were CB/EB. It was therefore not possible to compare those with/without visual experience. However, the results of Study 4 were consistent with Study 3 in that S reported a greater sense of Spatial presence, Engagement and Ecological validity when listening with SFX, compared to listening to dialogue alone. B&PS showed no significant difference between the two conditions. This is further evidence that S and B&PS process language differently. For S it seems that words are processed amodally and, in the light of findings from Study 3, prompt visual imagery but little more. By contrast SFX trigger more crossmodal associations, improving embodied presence. If, for B&PS, words are inherently modal (consistent with speech being processed in the visual cortex) this would account for the lack of difference they reported between the two conditions.

It was proposed in Chapter 1 (pp. 43 - 46) that words replace vision in integrating multisensory experiences for blind infants and that, even for B&PS adults, everyday experiences are accompanied by a greater use of explicit verbal information. Study 5, therefore, focussed on real world experience. Given the differences reported for Spatial presence and Ecological validity in Study 4, it concentrated on the impact of visual impairment on spatial mental models. In particular, it asked how sight characteristics affect the way spatial information is encoded, and what type of spatial information is stored by B&PS compared with S.

Study 5 compared descriptions of familiar routes recalled by 9 B&PS matched for age and gender with 9 S. 82% of the landmarks recalled by S were reported in the visual modality. Any congruent multimodal information was not reported suggesting it was encoded implicitly rather than explicitly. Details reported about the pathway were also visual (80%) and vague, rather than specific, as was information concerning distance between one landmark and the next. By contrast, B&PS gave very detailed descriptions. They reported largely multimodal information for landmarks (82%) and referred to significantly more small milestones along the pathway. Unlike S, B&PS also mentioned obstacles.

Metric information was highly specific and tended to use route- or body-based rather than objective measures. Interestingly, B&PS were less likely to report multimodal information for milestones (55%) than for landmarks (82%) even though milestones were detected by use of the long cane that provides both auditory and tactile/haptic cues. It was suggested that because such information can be directly experienced it is likely to be implicitly encoded, compared with distant landmarks that have to be encoded explicitly when a blind person is first familiarised with a route. It was noted that this type of experience, together with cortical plasticity, might account for B&PS demonstrating greater WM capacity for verbal information than S, such that verbal information carries less cognitive load.

Feist & Gentner (2007) have shown that, in sighted people, spatial scenes that are encoded with language affect later recall (participants were likely to exaggerate the spatial position of an object in relation to a building, choosing a picture of the object in a more extreme position than the one they were originally shown). This difference between implicit and explicit encoding, evident from Study 5, provides a possible explanation as to why, in a virtual environment, words can activate multimodal imagery in a blind person's SSM (spatial situation model), while providing less of a benefit to S.

9.2.2 Perception, experience and presence

Study 5 looked at the impact of visual impairment on real world experience to account for findings from Studies 2, 3 and 4. The main aim of these studies was to understand how visual impairment might impact on presence in mediated environments. Study 4 showed a difference between B&PS and S in response to a low-level immersive environment: an audio drama. Study 6 extended this to an audiovisual medium which B&PS access through a verbal commentary: Audio Description.

Prior to commencing the experimental studies, B&PS people's experience of AD was evaluated through interviews reported in Study 1. 16 people with a variety of sight characteristics described good and bad experiences of AD from which determinants of presence spontaneously emerged. Good AD was that which went unnoticed, just as awareness of media form must be suspended for an individual to feel present in any mediated environment. Yet AD simultaneously affects both media form (reducing a bimodal AV experience to a unimodal audio experience) *and* media content (by adding a verbal commentary that emphasises some visual elements at the expense of others, and by altering the "audio amalgam"). Participants suggested that the AD content may reduce presence by omitting salient information (leading to confusion or disengagement); providing conflicting information, or giving too much information thus adversely adding to cognitive load. Conversely good AD enhances presence by easing cognitive load, smoothing flow and establishing context by providing verbal details about location and characters. This can lead to greater Engagement, Spatial presence, Ecological validity and Social presence. User comments also reported some behavioural evidence of presence in the shape of physical responses to mediated content accessed via AD.

At the end of Chapter 3 it was suggested that AD replaces missing visual information with verbal *i.e. semantic* information. However, evidence from subsequent studies suggests that, for B&PS, verbal information is not purely abstract. Study 6 sought to clarify whether the addition of AD would increase or reduce emotion elicitation and thus alter presence. Furthermore, it compared the additional effect of the paralinguistic information carried in the vocal delivery of the human describer with an electronic voice conveying the same verbal content. 19 B&PS participants watched film clips known to elicit specific emotions in sighted people (either Fear or Sadness) with No AD, Text-to-Speech AD or Human Voice AD. They reported their responses using the shortened ITC-SOPI and selected items from the Elicited Emotions Scale (EES, Gross & Levenson, 1995). Physiological measurements of heart rate and electrodermal activity were also recorded.

Results were consistent with Study 4. Exposure to more of the soundtrack in the No AD condition did not lead to B&PS people reporting greater levels of presence. Nor did the soundtrack on its own elicit higher levels of the target emotion compared to listening with AD. However there was a difference depending on emotion categories. Film clips in the Sadness category satisfactorily elicited the target emotion with or without AD. The addition of TTS AD made little difference. The addition of HV AD eased Confusion but this, in itself, did not result in greater levels either of the target emotion or in dimensions of presence. By contrast, for clips in the Fear category, users could accurately *identify* the target emotion but they did not *experience* it directly without the addition of AD, and then only with HV AD. This was associated with significantly higher levels of Engagement, Ecological validity, fewer Negative effects and a trend towards greater Spatial presence.

Differences between the two emotion categories were explained by the amount of information available from the dialogue. Dialogue in the Sadness clips made them largely self-explanatory, whereas dialogue in the Fear clips was potentially misleading or omitted salient information. Differences were also explained in terms of Identification and Empathy with the characters. It was easier to identify with a character in a situation with which participants had personal experience (i.e. a bereaved character) than in a situation involving a serial killer or the supernatural. Stronger ratings of Identification and Empathy in the Sadness category were associated with higher levels of Ecological validity and Engagement and lower Negative effects, but not with Spatial presence.

The HV AD differed from the TTS AD in terms of its paralinguistic information i.e. the affective details conveyed through tone, pitch, pace and emphasis. This may explain why HV AD elicited Fear more successfully. However, it was also suggested that HV AD led to a greater sense of Social presence: sharing the experience with the human (if virtual) describer being more affective than sharing it with a synthesised voice. Although the intention was to compare purely semantic verbal information with the affective, paralinguistic element, HV and TTS AD also differed in that some participants reported the electronic speech to be more difficult to understand. The physiological data was inconsistent but HRV was significantly higher for Fear clips watched with TTS AD. Although this might indicate a heightened sense of Fear, it might equally reflect cognitive overload given the challenge of processing electronic speech. As mentioned in Chapter 1, Ahmed & de Fockert (2012) have found that attention is more easily distracted in conditions of high cognitive load and in individuals who have low working memory capacity. This is another explanation as to why TTS AD did not lead to higher levels of presence.

Perhaps the most interesting finding was that B&PS participants with visual experience showed no benefit from congruent bimodal cues e.g. when, in *Truly, Madly, Deeply*, the sound of Nina crying was supplemented by a verbal description of her tear-stained face. The quantitative data from Study 6 supports the comments of participants in Study 1 who disliked "redundant" information in the AD. It also supports the findings from Study 4: SFX that simply augmented dialogue did not increase blind people's experience of presence. It suggests that redundancy gains (whether from language and sound or from vision and sound) apply to S but not B&PS.

9.3 Implications of findings for models of AD

Chapter 2 presented 4 models of AD intended to illustrate the way AV media is experienced by people with different sight characteristics (CB/LB/PS/S). In the light of the experiments discussed above, those models are reviewed here.

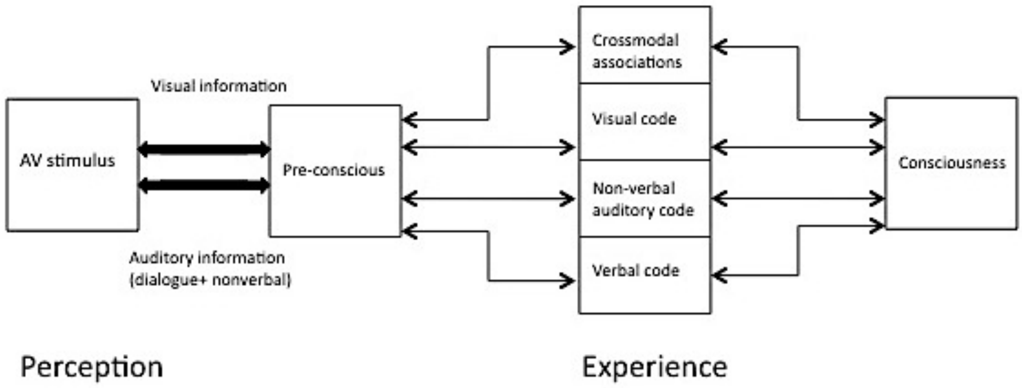


Fig. 9.1 Original model of an AV user with no sensory impairment

It was suggested (Fig. 9.1) that, for a sighted AV user, audiovisual information activates crossmodal and semantic associations, drawing on an individual's event-related knowledge base to produce a "virtual" multisensory experience. Such associations compensate for lack of direct input from other modalities. One modification, now suggested, is that the visual cortex actively suppresses information from other modalities, reducing input from the real world, so that presence in the virtual environment is more easily maintained. Auditory information from the AV stimulus is of increased importance when the visual stream is degraded or incongruent. In the revised model (Fig. 9.2) this is reflected by a dotted line between the visual and non-verbal auditory code indicating that the weighting of each may vary.

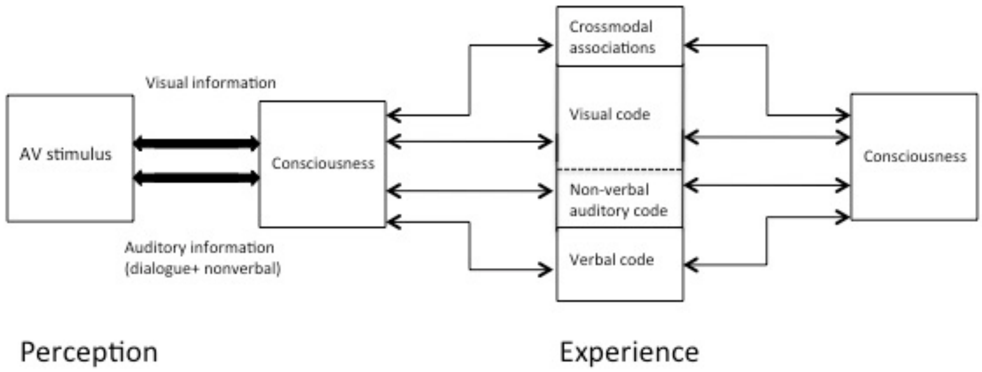


Fig. 9.2 Revised model of an AV user with no sensory impairment

The original model for an AV user with congenital blindness (Chapter 2, Fig. 2.6) showed simply an absence of visual information and visual code. The revised model (Fig. 9.3) suggests that, although the visual information is still absent, the missing visual code is replaced by an expanded verbal code. Crossmodal associations are mediated both by the verbal (dialogue and AD) and non-verbal auditory code. The shading reflects the finding that, in the absence of visual experience, crossmodal associations differ from those of other users.

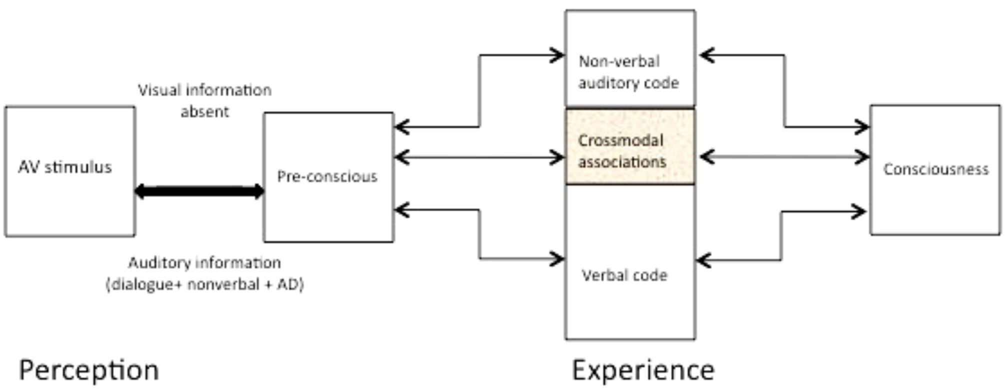


Fig. 9.3 Revised model of an AV user with congenital blindness (CB) watching with AD

By contrast, for at least some LB users, a residual visual code is likely to maintain crossmodal associations. These can equally be triggered by verbal (dialogue or AD) or non-verbal sound (Fig. 9.4).

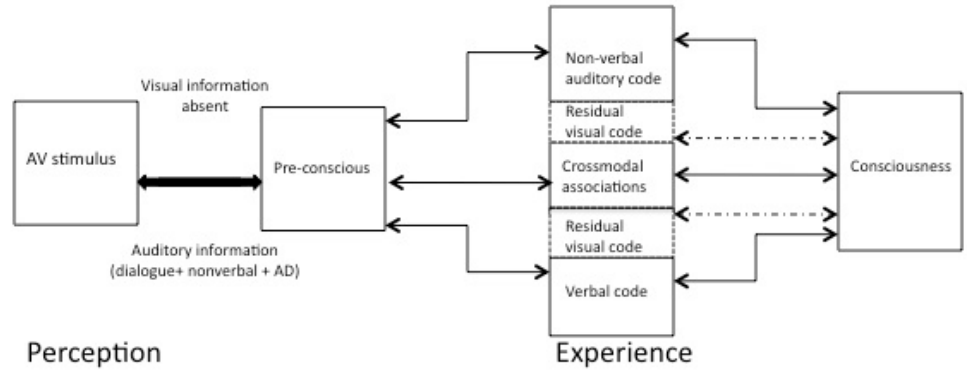


Fig. 9.4 Revised model of an AV user with late-onset blindness (LB) watching with AD

A user with partial sight (Fig. 9.5) will have access to intermittent or degraded visual information. Either, then, they have direct access to a visual code or, when visual information is unreliable, are able to fall back on a residual visual code, prompted by non-verbal or verbal sound. However, sound associations may differ from those of S through reduced exposure to environmental regularities.

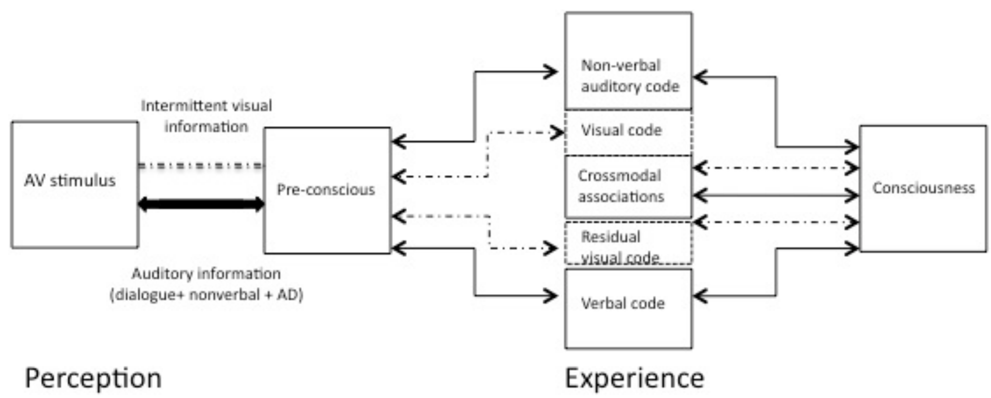


Fig. 9.5 Revised model of an AV user with partial sight (PS) watching with AD

In addition to theoretical implications, the studies presented here have practical applications. Study 2 suggests that sound associations, thought of as "universal", cannot be taken for granted in those with impaired sight. It may be useful for teachers to be aware that such associations need to be made explicit to aid CB children, whose language development is often delayed, in catching up with their sighted peers. It may also affect word choices made by audio describers who are currently encouraged to use language techniques such as onomatopoeia, for example choosing short, staccato words such as "jab" and "thwack" when describing a fast movement sequence such as a fight (Fryer, 2009). In addition to AD implications discussed in the context of each study, the relevance to AD of Gibson's ecological approach to visual perception, raised in Chapter 1, is presented in Appendix D. Suggestions for future research in AD are proposed below, in Section 9.6.2.

9.4 Implications of findings for models of imagery and perception

In discussing spatial imagery, Struiksma et al. (2009) suggest that the image is maintained by information from different modalities (weighted in accordance with their importance) plus amodal information from spatial language. As discussed in Chapter 1, they place language in opposition to perceptual experience. They suggest that when going to buy a pineapple one can either conjure up a multisensory mental image of it or act on a set of abstract, verbal instructions. The findings of this thesis suggest that the word "pineapple" will trigger a largely visual image in the mind of a sighted person. Yet if the image were initiated with the smell or taste of pineapple, a richer, multisensory image would be produced. Struiksma and her team argue that, for people with impaired vision, more information from non-visual modalities compensates for reduced visual information. The logical extension is that a blind person would be able to imagine the smell of a pineapple more strongly and more vividly conjure up its spikiness, its weight and shape in the hands. Yet evidence from the studies reported here suggest that representation will not be richer or *more* vivid than the multisensory image of a sighted person. The difference is that the image will be as rich if triggered by a verbal cue as by a sensory one.

The theoretical issue here is whether language is modal or amodal, and consequently, whether the image itself is modal or supramodal. It is possible that language can be both: modal when sensory information is explicitly encoded; amodal when not tied to a specific sensory experience. In the former situation, language may affect the image by explicitly increasing awareness of the contribution of non-visual modalities. As explicit encoding is routine for B&PS this accounts for a multisensory image being triggered by words. In the absence of (dominant) vision, then, the contributions of non-visual modalities are revealed.

If this is the case not only for imagery but for perceptual processing in general, it is in line with a recent suggestion by Kupers and Ptito (2013) that, rather than cortical reorganisation, the “blind” brain is simply strengthening existing neuronal connections. They cite evidence from animal studies that show, in normally developing individuals, non-visual inputs are routed via the occipital cortex. These inputs are masked (or arguably inhibited) by visual processing. In the absence of visual processing, areas of cortex thought to be visuo-specific can be recruited by other sensory inputs, and beyond this, by language. This is not restricted to people with a visual impairment. Pascual-Leone and Hamilton (2001) showed that five days of blindfolding is sufficient for sighted participants to begin to process auditory and tactile information in their visual cortex.

In general, there is a move away from a model of discrete modal areas in the brain in favour of one in which cortical recruitment is task-based rather than modality-based (James et al., 2011). The evidence from the studies presented here support that view, putting forward the case that, rather than enhanced non-visual sensory modalities, it is language that compensates for lack of sight in the blind.

9. 5 Revisiting the questions

Taken as a whole, the studies in this thesis support the hypothesis that similar performance between CB/EB/LB/PS results from different processing systems. Given the small number of CB/EB participants, it is difficult to draw definitive conclusions. However, Study 2 (Bouba-Kiki) suggests that lack of visual experience leads to differences in crossmodal connections in non-visual modalities. This was also evident in differences in mental imagery in Study 3. Yet differences in processing had no apparent effect on outcomes in Studies 4, 5 and 6. A verbal encoding strategy produced similar responses from LB/PS and CB/EB in Study 5, and in Studies 4 and 6 verbal information had a similar effect on levels of presence and emotion elicitation, whether or not participants were able to access a visual referent.

There was little evidence to support the idea of auditory compensation in mediated environments. Music and SFX alone did not induce presence, nor elicit target emotions. In the absence of coherent (narrative) information from the dialogue, the addition of verbal information significantly improved understanding. When delivered by a human voice, AD went beyond that, raising levels of presence and empathy, provoking an emotional response in visually impaired participants and allowing them to experience the mediated environment as real. AD did not lead to greater levels of presence in LB/PS than in CB/EB. Visually impaired participants benefitted equally in terms of accessing the visual stream of AV content through the medium of words.

9.6 Directions for future research

9.6.1 Imagery

The studies presented here relied on small samples and replications would be welcome. They also raise a number of questions for future research both with sighted people and people with a visual impairment. Study 2 implies that crossmodal correspondences between non-visual modalities can be affected by the strength and vividness of visual imagery. This is known to vary between individuals. Paivio (1977), for example, developed his “Visualiser–Verbaliser Questionnaire” to assess preferred cognitive style. This would be a useful starting point to explore the indirect role of visual imagery in aiding crossmodal integration in non-visual modalities.

Neuroimaging could also provide a way of measuring activity in the visual cortex during direct perception of non-visual stimuli. For example, Bidet-Caulet, Voisin, Bertrand and Fonlupt (2005) showed that, for S, listening to footsteps (but not meaningless sound) activated the posterior superior temporal sulcus (STSp). As this area of the brain is also activated when watching bodily movement (but not movement of objects) it suggests that the sound directly prompts visual (and motor) imagery. The sound of footsteps also activated frontal and parietal areas associated with visual attention. This is in line with the verbal reports of visual imagery stimulated by SFX in Study 3. It would be useful to see if imagery as reported for other non-visual modalities was similarly backed up by neuroimaging data.

Study 3 showed limitations with existing auditory imagery scales. Zatorre, Halpern and Bouffard (e.g. 2010) have run a number of studies using the Bucknell Auditory Imagery Questionnaire without encountering the ceiling effect shown by the CAIS. This could be used to test whether the lack of evidence of enhanced auditory imagery in B&PS shown in Study 3 was the result of flaws in the measure chosen. Again it would be useful to compare neural correlates in people with varying sight characteristics.

9.6.2 Audio description

One of the main findings for AD is that information sighted people view as reinforcement can be regarded by AD users as redundant. It is important to test the extent to which aspects of appearance can be removed from a description before it loses its efficacy. Study 6 could be repeated comparing the standard AD script with a minimal version, removing all references to specifically visual qualities (e.g. references to colour such as Bambi's *brown* body, the *green* blades of grass and *white* snow). It would also be interesting to substitute these with qualities associated with non-visual modalities such as temperature, odour or shape to see whether that enhanced presence.

Another finding was the type of embodied imagery favoured by B&PS participants. A comparison of descriptions given in the first person and the third person perspective may show the former to be more vivid, giving rise to greater presence than the latter. Similarly, in a gallery displaying furniture, for example, it may be more evocative to describe how it feels to sit in a particular chair than to describe its appearance.

Given the findings about prosody, more research could be carried out into the impact of AD delivery. Age, gender or social background of the describer, for example, might effect engagement with the source material. Some ADs have been recorded with celebrity voices. It would be interesting to see whether this aids or reduces Social presence and/or Engagement.

Study 6 also suggested that physiological responses to AD might depend on cognitive load. Complexity of the AD script, not only in terms of density, but also lexical difficulty is another area of potential research. Guidelines in some countries, such as Spain, advise that language should be "simple, with directly constructed phrases" (Rai et al., 2010, p.10). This could provide a baseline for comparison, taking into account verbal ability of B&PS participants. This might be particularly useful with reference to films or programmes for visually impaired children who are known to have delayed speech development.

In addition to AD *content*, there is potential for exploring AD *form* in terms of technical delivery. How is the reception of AD affected by delivery via headphones compared to broadcast through speakers? Potentially headphones are more immersive yet reduce access to the social side of experiencing the programme or film in the company of others. Could this be improved by feeding in ambient sound from the auditorium? Could AD delivered by a surround-sound system enhance a sense of Spatial presence, with the direction of the AD voice equating to the positioning of events on stage or screen? Does multi-channel sound alter response to content? Even basic considerations such as the balance of AD and source sound in the mix offer avenues to explore.

There is also potential for researching the effects of AD on other would-be users. For example, prosopagnosics (see Appendix A) find it difficult to recognise characters from one scene to the next. They might benefit from AD. Potential educational benefits include guided attention, repetition or illumination of complex linguistic terms. This could be studied with sighted children (see Krejtz et al., 2012) or with learners of English as a foreign language.

9.6.3 Orientation information

The RNIB is currently exploring ways of aiding B&PS people in wayfinding. One suggestion is to provide a “help me” service, where a (presumably sighted) buddy would provide information about a location. This relies on the “buddy” knowing what type of information is required. Sighted "buddies" should be trained in making details explicit through verbalisation. The same is true for those who provide orientation information for audio guides at museum and heritage sites, or for curators or guides who may be asked for directions face-to-face. The suggestions of Study 5 (that highlighting small-scale milestones may help blind people keep on track between major landmarks on their journey and preference should be given to information that can be accessed multimodally) could be tested experimentally. The design of Study 5 could also be applied to indoor as opposed to outdoor spaces, to compare differences in navigation strategy when conditions such as atmosphere and temperature are more controlled.

9.6.4 Presence

Having shown the benefits of presence as a measure of user experience for people with a visual impairment, it would be valuable to explore presence in audiences with other sensory disabilities. Wissmath, Wiebel and Groner (2009) have used presence as a measure to compare subtitled and dubbed speech for hearing audiences in Switzerland watching English-language imports. It would be possible to carry out similar research looking at the effect of caption styles on deaf-and-hard-of-hearing audiences. It would also be interesting to compare, for example, different positions of display boxes for live captioning in the theatre.

9.7 Conclusion

This thesis makes an important contribution to presence research. It provides qualitative and quantitative evidence that people with a major sensory impairment can experience presence in audiovisual environments. This may be direct, through the dialogue, music and sound effects of the source material, or indirect via an additional verbal commentary: Audio Description. Presence can be attained whether or not the AD user has visual experience. For all people with a visual impairment words trigger a multisensory experience. This finding informs the debate on multisensory processing, countering suggestions that sight is essential to integrate information from different sensory modalities. The studies here put forward a case for linguistic, rather than sensory, compensation for those with a visual impairment. Findings support the theory of language as multimodal, resulting from explicit encoding of sensory experience. Furthermore they suggest that language can replace vision as the framework for integrating sensory experience. This thesis also makes a practical contribution to AD research. It shows what type of verbal information is required and what may be regarded as redundant. By demonstrating that presence can be a useful measure of AD quality, it points the way towards improving an important means of access to audiovisual media for people with a visual impairment.

Appendix A

A1. Document produced for the International Telecommunication Union (ITU) Standards Committee on Access

Audio Description and Audio Subtitling: Who Needs It?

AD User Personas

Audio Description (AD) and Audio Subtitling (AS) provide access to audiovisual media for people who may otherwise be excluded. What is striking about the audience for these services is its sheer diversity. To illustrate this, here are 7 "personas". None of these people are real. They are composite characters, based on research interviews with actual users of AD and AS in the UK. In addition 6 personas have been created representing potential users of AD who are sighted but have particular perceptual, cognitive and/or communication difficulties. They are representative of the potential AD and AS audience across EU member states.

CURRENT AD USERS

Persona 1: John, aged 34 (customer service adviser) John lives in West London, in a one-bedroom flat in sheltered accommodation. John is very independent. He's a guide dog owner and travels to work in central London by tube. Some evenings John might stay in and watch TV or DVDs. More often he meets friends after work and goes to the cinema or the theatre. John has been blind since birth. He uses AD whenever possible and finds it crucial for following the plot. John has no visual memory and is not interested in some of the details audio describers include, such as style of dress, but he accepts other AD users ("women!") might want to know about this. He is keen to know about the physical appearance of a character or location but only if it is relevant to the storyline. His favourite programmes include crime thrillers and American drama serials such as Lost. John finds AD “particularly helpful in the spaces where there is very little dialogue. Films can be boring and I give up if there are lots of spaces. The main fault with a lot of AD is that they don't explain enough who's speaking. It can take me half an hour to recognise a voice”.

Persona 2: Peggy, aged 72 (retired) Peggy lives with her (sighted) husband. They do not go out to the theatre or the cinema much, but they often watch TV. Peggy was diagnosed with retinitis pigmentosa 8 years ago and her sight has been gradually getting worse. Her favourite programmes include soap operas and quiz programmes. Soaps she was able to watch before she had access to AD because the characters and locations were familiar, but she finds AD gives an added extra. In Coronation Street, Sophie and Sian are due to get married and Peggy is looking forward to hearing a description of their wedding dresses. Before AD she was reliant on her husband telling her what was happening. Peggy says “He’s not very keen on Corrie. The big bonus with AD is that I can just listen to it without having to nudge him to ask what's going on”.

Persona 3: Sandy, aged 41 (teacher) Sandy is a primary school teacher. She lives in a village in Hertfordshire and walks to work. Sandy has nystagmus and her vision can be very variable, depending on lighting conditions and how stressed or tired she is. For much of the time, Sandy doesn’t always bother using AD. She likes news and current affairs programmes (which are rarely described). She also enjoys documentaries. As her biggest difficulty is reading text on screen, AD comes in handy when a person is speaking in a foreign language, as the describer voices the subtitled speech. Sandy says she also uses AD for certain programmes, such as dramas with atmospheric lighting. She says "I can be really lost without AD especially if it's, say, a film set in Victorian times and very dark, backlit. Horrible! Then I find myself enjoying other bits of the description, such as facial expression. I realise how much I'm missing".

Personae 4 & 5: Michael, aged 28 (musician) and Hanna, aged 27 (full-time mum) Michael and Hanna are both partially sighted. They met at the Royal National College for the Blind, where they were at school. These days they live on the outskirts of Bristol. They have a daughter, Lily, aged 6, who is sighted and attends the local primary school. Michael used to play rugby. He goes to matches when he can, or watches them on TV. Sports programmes only rarely have AD, so Michael relies on the radio commentary instead, which is more descriptive than the TV commentary. The family keeps the AD switched on all the time so they don't miss anything with description. Michael watches medical dramas like Casualty, which has its drawbacks. "If you're eating tea and there's a gory operation going on, sighted people can shut their eyes. If I shut my eyes, I still hear about the blood and stuff from the AD!" Hanna says, "When Lily was little, we would sit down and watch kids' programmes like Peppa Pig[8] together. It was really important for me to know what was happening so we could talk about it. Sometimes Lily complains about the AD but mostly she's used to it. She's

[8] Peppa Pig is a British animated television series created directed and produced by Astley Baker Davies and distributed by eOne Entertainment. To date, three series have been aired. It has been shown in 180 territories.

coming with me to the pantomime. It's being audio described and Lily asked if it was going to be 'one of mummy's special shows'".

Persona 6: Bob, aged 79 (retired engineer) Since suffering a stroke 3 years ago, Bob has lived with his son, Phil, Phil's wife Erica and their 3 teenaged children. The stroke caused Bob to lose his sight completely. He finds AD allows him to watch TV with the family, so they can all sit in the same room in the evenings. Bob listens through headphones, so the others don't have to hear the AD. Bob finds AD "fills in the gaps", so "I can follow what's happening". He finds AD of comedy programmes especially useful so he doesn't have to ask what the others are laughing at. It can be particularly embarrassing if the comedy is a bit risqué. "When the describer for Shameless talked about "a couple at it on the pan,"' Bob said, "I was in stitches. It was great that we could all laugh together, without my grandsons having to explain the joke. If we're watching a serial, sometimes Phil asks if we've seen such-and-such a character before, and I can say 'yes, it's Lord Haversham' because they've told me in the audio description".

Persona 7: Kieran, aged 4 Kieran was born with anophthalmia. His parents, Bill and Alison, devote a lot of time to a charity working with families of children who are born blind. Such children can be slow to develop language skills compared to their sighted peers. Bill and Alison feel that watching children's TV with AD helps Kieran pick up new words and phrases. Alison says, "Kieran loves watching the Octonauts[9], and sometimes pretends he's one of the

[9] The Octonauts preschool animated series (produced by Silvergate Media / Brown Bag Films) is currently airing on CBeebies (UK) and in many other territories across the globe.

characters. If I'm in the kitchen getting his tea ready, he'll tell me what's happening in the latest episode. And it means he can join in, when his friends pretend to be Octonauts at playgroup."

POTENTIAL AD USERS

In the UK, research into AD use has tended to focus on its potential for people (and the families of people) with sight problems. However AD and AS may have broader appeal. The following personas may be best thought of as potential users of AD and AS.

Persona 8: Jake, aged 11 (attention deficit disorder) Jake was cross when he came home from school, because his teacher had been mean to him. She told Jake off for fidgeting when the class was watching a cartoon about the human body and the circulation of the blood. Jake couldn't answer the questions the teacher asked at the end of the cartoon. Jake has difficulty concentrating and gets bored easily. He says there are so many things to look at, how's he supposed to know what's important? With AD, Jake's attention is directed to the key events on screen and strange names ("corpuscles" "capillaries") are repeated by the describer, making them easier to remember.

Persona 9: Sonali, aged 24 (English as a 2nd language) Sonali came to England with her husband Nikhil, 6 months ago. Sonali sees a lot of Nikhil's family with whom she can speak her native language. Sonali, who thinks of herself as a very confident person, was hesitant about going out in case she did not know what people were saying. Her cousin told her about AD. Sonali finds AD is helping her pick up more English words and phrases. Watching the pictures makes it easier to understand the description. AD also helps with English idioms used in the dialogue. Sonali says, “On Eastenders, Phil asked if he could ‘take a butcher’s’. The describer said ‘Phil looks at the photos’ and I knew that this must be a funny way of asking to have a look at something! Sometimes I watch Bollywood movies with audio description. The describer reads out the English subtitles, so I can see the written words and hear how you’re supposed to say them”.

Persona 10: Alice, aged 83 (early stages of Alzheimer’s) Heather has been looking after her mother, Alice, for the past 6 months. Alice is starting to have difficulties with her short-term memory and her hearing is not as good as it used to be. Heather works full time but she lives just around the corner, and calls in every morning to help Alice get washed and dressed and have breakfast. Heather calls in again after work to prepare her mother’s supper and to get her ready for bed. Friends sometimes drop by during the day, but Heather was worried that her mother wasn’t getting enough mental stimulation. “Sometimes we would watch TV together in the evening and Mum would just drift off into her own world or talk about something completely unrelated, something that had happened to her years ago. With AD, she doesn’t get so muddled and can follow the programme through to the end. We can even have a chat about it afterwards.” Alice says, “With AD, it feels as though the characters are talking more slowly so I can take in what they’re saying. And I find it easier to follow the story.”

Persona 11: Ella, aged 3 (early years language acquisition) Ella is nearly 3 years old. Ella is sighted but when she plays with the TV remote, sometimes she accidentally turns on the AD. Ella does not usually notice, but Ella's mum, Rachel, knows when it has happened. "Ella's a real parrot," Rachel says. "She'd been watching Charlie and Lola on CBeebies and they went to the zoo. Ella told me 'Lola saw an elephant. It was big, grey and wrinkly, with smooth, white tusks.' I thought how does she know words like 'wrinkly' and 'tusks' and I realised she must have copied them from the audio description."

Persona 12: Sam, aged 17 (media student) Sam is at Sixth Form College. His favourite subject is Media & Film Studies. Sam has started watching films with AD. He says "the description points out things I haven't noticed. Sometimes there are special ADs that mention the cinematic elements of film - the different cuts and dissolves, and whether it's a close shot or a zoom or a tracking shot. That's really brilliant. The describer deconstructs the film for me and I can work out how the director's manipulating the audience by using visual effects. And now I've got used to listening to the AD, I can go and get myself a drink, or something, and still hear what's going on, so I don't miss anything".

Persona 13: Jacob, aged 43 (psychologist) Jacob is a highly qualified, well-respected psychologist. Recently, following a stroke, he has developed problems recognising faces, even those of his wife and daughter. He has also ceased to recognise what were once familiar landmarks, although he has no difficulty once they have been identified by name. Similarly, Jacob can recognise his family by the sounds of their voices and by the way they move. Jacob is not severely debilitated by his condition, which is called prosopagnosia or ‘face-blindness’, but it can lead to acquaintances thinking him rude for ‘pretending’ he has never met them before. Prosopagnosia also makes it hard to follow a drama or a film. Jacob says “by supplying the verbal labels that are part of my daily coping strategy, AD is extremely useful. Prosopagnosia often goes unrecognised, but is thought to affect between 1 – 2.5% of the population. If more prosopagnosics knew about AD, I’m sure they too would find it beneficial”.

Appendix B

B1: Study 2

Gender	Age	Age registered blind or partially sighted	Visual Acuity 1 = profound sight loss 7 = fully sighted
F	30	birth	1
M	27	birth	1
M	59	1	1
F	38	5	1
M	33	5	1
M	58	10	1
M	30	5	2
F	30	7	1
F	43	11	2
M	38	13	1
M	59	31	2
F	42	28	1
M	44	30	2
M	64	50	1
M	58	54	1
F	78	76	2
M	52	15	5
M	45	17	3
F	44	21	6
M	41	20	3
F	52	36	2
M	40	30	5
M	56	43	3
M	80	78	4
M	59	47	5
F	64	52	5
M	61	50	5
F	63	53	5
F	47	40	4
F	49	birth	5
M	48	11	2
M	30	12	2
M	24	birth	1
M	65	birth	1
M	60	birth	1
M	45	4	1
F	49	21	1
F	52	39	5
F	30	birth	3
F	42	7	6
F	77	16	1
F	48	5	5

Table B1 B&PS participants

B2 Kiki and Bouba: Script for Procedure

I have 4 bags. In each bag there are 2 objects. I will hand you the bags, one at a time. I want you not to look inside, but to feel inside. [hands over the 1st bag].

Put your hand in here. You might want to use both hands. Now, one of these objects is called Kiki, and one of these objects is called Bouba. I would like you to bring out Bouba [Kiki] for me.

Thank you. Put Bouba [Kiki] back in the bag. Here’s the 2nd bag. Inside you’ll find Kiki and Bouba again. This time I’d like you to bring out Kiki [Bouba] for me.

Thank you. Put Kiki [Bouba] back in the bag. Here’s the 3rd bag. Inside you’ll find Kiki and Bouba again. This time I’d like you to bring out Bouba [Kiki] for me.

Thank you. Put Bouba [Kiki] back in the bag. Here’s the final bag. Inside you’ll find Kiki and Bouba again. This time I’d like you to bring out Kiki [Bouba] for me.

Thank you. Finally I’d like you to tell me why you made those choices.

Appendix C

C1: Study 6 Participants

ID	Gender	Age	Age registered Blind or Partially Sighted	Visual Acuity* (1 – 7)	Aetiology
E01	M	48	11	2	autosomal recessive retinal dystrophy
E02	M	30	12	2	optic atrophy
E03	M	34	4	1	persistent hyperplastic primary vitreous
E04	M	24	birth	1	anophthalmia
E05	M	65	birth	1	retrolental fibroplasia
E06	F	64	50	3	retinitis pigmentosa
E07	M	45	16	2	retinitis pigmentosa
E08	M	46	36	5	macular dystrophy
E09	F	59	37	2	congenital glaucoma (PS from birth)
E10	M	60	birth	1	rubella
E11	M	45	4	1	glaucoma with retinal detachment
E12	F	49	21	2	glaucoma
E13	F	44	11	1	shrunken eye syndrome
E14	F	52	39	5	retinopathy and glaucoma
E15	F	30	birth	3	underdeveloped macular; congenital nystagmus
E16	M	60	birth	1	retinal blastoma
E17	F	42	7	6	double hydrocephalus (PS)
E18	F	77	16	1	retinitis pigmentosa (PS from birth)
E19	F	48	5	5	optic atrophy

Table C1 B&PS Participants
*Visual acuity coding (after Douglas et al., 2006). Which of these best describes your sight with glasses or contact lenses if you normally use them? 1 = I have no light perception; 2 = I can tell by the light where the windows are; 3 = I can see the shapes of furniture in the room. I can recognise a friend by sight alone if: 4 = I'm close to their face; 5 = I'm at arms' length away; 6 = I'm on the other side of the room; 7 = I'm on the other side of the street.

C2 Study 6 Documentation

C2.1 Sample Consent Form (font size reduced for purposes of thesis)

Project Title: Emotion Elicitation through AD

Researcher: Louise Fryer **Supervisor**: Dr. J. Freeman

- This research is part of a PhD project at Goldsmiths College, University of London. Thank you for agreeing to take part.
- The session is expected to take about 75 mins. Your heart rate and skin conductance levels will be monitored throughout the session. This requires the use of light-weight, non-invasive equipment. You are welcome to examine the equipment, or try it out, before committing yourself to taking part. Once the session has begun, you may ask to stop at any time.
- You will be asked to watch 6 short film clips. Please remain as still as possible while watching. The clips have a range of classifications from U to 18. Some come from thrillers and some involve scenes relating to bereavement but there is no graphic sex or violent content. If you find any clip upsetting, remember, you may ask to stop whenever you like.
- After each clip, you will be asked a series of questions. If there are any that you do not wish to answer, please say so.
- We will also ask some more general questions about you, to enable us to analyse your feedback in context. All information you provide will be treated confidentially.
- Data from this session may be used in reports related to this project. If you have any questions, please ask the researcher before you complete this form.
- If you are still willing to participate, please initial the following statements, and sign and date the following sections:

1 I confirm that I have read and understood the information given for this research or have had the information read to me, and have had the opportunity to ask questions.

2 I consent to take part in this research session.

Name of Participant	Date	Signature

C2.2 Introductory script

You are about to watch clips from 6 different films. Each clip lasts around 3 minutes, though some are slightly shorter and others a little longer. Some have audio description and some do not. Before each clip starts, there will be a blank screen and silence for 30 seconds. I would like you to sit as still as possible during this time, during the clip itself, and again for 30 seconds once the clip finishes. I'll let you know when that 30 seconds is up. At that point I will place the slider on your lap and ask you a series of questions. I'd like you to respond by using the slider just as you did earlier, except for a few questions which require a verbal answer. When you have completed all the questions about that clip, I'll ask you to sit still again as you wait for the next clip to begin. If you have any questions please ask them now. If you are happy to begin, I will start the session.

C2.3 Question Sheet

A: I am going to read you a list of 6 emotions. The first is Arousal, by which we mean alert/tense/excited/ready at the high end, and tired/calm/bored/inactive at the other. Where moving the slider to the extreme left means you did not feel even the slightest bit of the emotion, and to the extreme right means the most you have ever felt in your life, please position the slider to reflect greatest amount of that emotion you felt at any time during the clip. The emotions are Arousal; Confusion; Interest; Fear; Sadness; Tension

Now, where the extreme left corresponds to “strongly disagree”; and the extreme right to “strongly agree” please position the slider in response to the following statement:

B1: I felt myself being drawn in
B3: I lost track of time
B5: The scenes seemed natural
B9: I felt I was visiting the places in the scenes
B11: The content seemed believable to me
B15: I felt the environments in the clip were part of the real world
B17: I paid more attention to the scenes than to my own thoughts (e.g. personal preoccupations, etc.)
B18: During the clip I had a sense of “being in the scenes”
B25: I felt surrounded by the scenes
N 1: I experienced sensations such as dizziness, disorientation, nausea, a headache, or tiredness

C1: How much did you identify with the concerns of one or more of the characters – i.e. to what extent did the clip address issues relevant to you? The extreme left corresponds to “not at all” and the extreme right to “very strongly”

C2: How much did you empathise with the characters – that is feel the same emotions they were experiencing? The extreme left corresponds to “not at all” and the extreme right to “very strongly”

C3: What emotion do you think the clip was trying to make you feel?

C4: Have you seen this film before? 0 = No; 1 = Yes without AD; 2 = Yes with AD

You are about to watch another film clip but first take a moment to clear your mind of all thoughts, feelings and memories.

C3 AD Scripts

C3.1 Bambi Dur: 03.05

00.00 Music
The little deer, Bambi, shivers in a thicket

00:15 ...but it won't last forever
He snuggles down next to his mother

00:19 ...yes, I know
He nestles against her. On the white meadow, a few tufts of grass sprout up through the snow

00:33 ...Bambi come here
The little deer bounds over *["look"]* and pricks up his ears

00:40 ...new spring grass
Bambi nibbles it. His mother eats alongside him, their brown bodies dark against an unbroken expanse of white snow on the edge of the forest.

00:51 ...music
Bambi tugs eagerly at the tiny green blades. His mother raises her head, alert. She looks around, ears twitching

01:05 ...the thicket
They race for cover leaping an icy stream

01.22 ...keep running, keep running [Bang!]
Bambi skitters ahead, ears flattened, zig-zagging towards their winter shelter – a dry space under the branches

01:40 ...we made it mother. Mother?'
He pads back out.

01.44 …Mother'

Snowflakes are falling, almost blotting out the trees

01:53 …Mother where are you?'

Bambi trots this way and that

02:00 …Mother!

He pauses every few steps to peer through the snow

02.07 *…Mother, Mother!*

The light is fading fast

02.19 …music

Bambi reaches the edge of the trees and falters

02:29 [gasp]

A huge stag looms

02:34 …with you any more'

Bambi stares at the stag. The fawn's big round eyes fill with tears and he lowers his head sadly. The stag stands proud. Slowly, Bambi looks up at him.

02:54 …Come, my son'

The stag stalks away. Little Bambi trails behind him, stops and looks back with a sorrowful glance. The stag waits without turning. Bambi follows him on.

C3.2 Four Weddings and a Funeral Dur: 05.20

00:01 …music

A housing estate with the bridge of the Dartford Crossing rising up beyond the rooftops against a cold grey sky.

00.11 ...music

Mourners with umbrellas gather around a funeral cortege. Matthew ushers Gareth's parents towards the waiting limousine. He helps them into the back seat and joins them.

00.31 ...music

The car tails the hearse down the rain-spattered road. Mourners follow on foot. Charles and Scarlett hurry up the path to the church

00.40 ...music

Inside, Charles makes his way between crowded pews, glancing at Carrie as he passes. She looks away. Charles raises a hand in an awkward greeting and squeezes in beside Fiona. She wears a black beret, her face sombre. She passes a handkerchief to David's girlfriend in the pew behind. Charles puts on his glasses and slips a comforting arm around Fiona's shoulders. The vicar steps up to the lectern.

01.22 ...to say a few words

Matthew gets up from the pew at the front of the Church. Gareth's parents watch with anguished faces as Matthew stands at the head of the coffin.

02.20 ...strange experimental cooking

Gareth's mother gives a sad smile

02.39 ...highly vocal drunkenness

An older man laughs – and checks himself

02.44 ...remember him

Matthew swallows

03.00 ... big hearted

Members of the congregation nod

03.20 … what I thought about him
Matthew's face is pale, eyes dark-ringed

03.42… actually what I want to say
He reads

04.47 …I was wrong
The coffin's carried to the hearse

05.00. …dismantle the sun
Bewildered, Gareth's parents look on

05.11 …to any good
Matthew stands by the pallbearers [*ignition noise*]. They move off with the hearse. Matthew turns to Gareth's parents and they fall in behind, following the sleek black limo.

C3.3 Truly, Madly, Deeply Dur: 03.15

00.01 Music
Nina pegs out the washing. She glances up at Alex and the builders watching her from the kitchen window, and smiles. They're washing up.

00.30…couldn't get the tube down
Nina, in her 30s, looks thoughtful, the wind tousling her hair as she pegs a shirt. It flaps in the breeze

00.50 …I'm completely numb
Nina's with her therapist, head in hands, cheeks tear-stained

01.14 …it's ridiculous
A teardrop trembles on the tip of her nose

01.26 …shouldn't do this

She covers her face with her hands

01.29 …[sniff]

A mousey woman watches dispassionately

02.04 …so angry with him. I'm so angry with him.

Nina bows her head, shaking

02.13 …[sniff]

Tears seep through her fingers

02.25 …[sniff]

She rubs at her eyes, wipes her nose with the back of her hand.

02.37 …[sniff]

The therapist reaches forwards calmly and turns off a tape recorder.

02.50 …[sniff]

Nina's face is blotchy and red

02.56 …[sniff]

The woman gives a small tight smile

03.08 …can I?

Nina takes a tissue *[sniff]* and goes.

C3.4 Silence of the Lambs Dur 03.30

00:01 ...music

An old mobile home beside train tracks under a dismal sky. Across the tracks an isolated house flanked by bleak winter trees. A car parked at the end of the path. Inside, the female FBI agent walks through the hall towards Jack.

00.54 ...don't seem to have the first clue

Starling surveys her surroundings. A death's head moth lands on bobbins of coloured thread. Starling swallows, a glint of fear in her eyes.

01.10 ...like that? No

Jack nods

01.14 ...nothing like that

Starling unclicks the safety lock on her revolver. Jack's sorting papers. In the adjoining kitchen, a gun lies on the hob of a grease-encrusted cooker.

01.24 ...here's that number

He holds out a card

01.30 ...use your phone please?

He smothers a laugh

01.34 ...chuckle

Starling's gaze briefly falters

01.48 ...spread your legs.

Jack lifts his hands casually

01.53…freeze!

He turns and slips through to the kitchen. Starling follows, edging round the doorframe. The room's empty. She shrugs off her jacket. Pulls open a door, kicking it as it tries to swing shut.

02.06 ...music

Steps lead down to the basement. Squatting on the top step, she points her gun in shaking hands. In the shadows below, a rusty drier and a silk scarf on a low cupboard between two closed doors. She inches down the stairs.

02.29 ...music

Eyes wide, face tense and frightened, she approaches the doors. Choosing the nearest she sidles up to it.

02.51 ...music

The door opens towards us, Starling's anxious face framed in the light beyond. She ventures forwards.

03.05 ...music

Light filters through a small window, at the end of a brick corridor. Pinned to another closed door, a map of the States, ripped around Michigan.

03.14 ...music

Starling narrows her lips and puts her shoulder to the door.

03.20 ...music

Back to the wall, she gazes down at a dismembered body, putrefying in a bath. Darkness.

C3.5 The Shining Dur: 02.15

00.01 ...music

The hotel, closed for winter, is shrouded in snow. Inside, little Danny lines up toy cars using the geometric patterns of the carpet that lines the corridor as a make-believe roadway. A tennis ball rolls towards him. He looks up. The corridor is empty.

00.30 ...music

Danny gets to his feet.

00.35...music

His breathing quickens

00.43 ...music

He steps over his cars and pads forwards, eyes wide, face solemn, an Apollo rocket pictured on the front of his hand-knitted jumper.

00.55 ...music

The corridor stretches away from him. The doors to all the rooms are closed… except one.

01.05 ...music

As Danny draws nearer to room 237, the door's ajar, the fob swinging from the key in the lock. Mirrored cupboard doors inside reflect a glimpse of the lit interior.

01.22 ...music

The double bed is made up. Lights gleam through the half-open door to the ensuite bathroom

01.33 ...music

As we approach, the door blocks a clear view.

01.41…music

A hand pushes it slowly open.

01.47 …music

The bath's at the far end, the shower curtain drawn partway across. Through the translucent plastic, the shadow of a figure. Danny's Dad, Jack, stands in the doorway, staring towards the bath. He swallows nervously.

02.04 …music

In the bath, a hand reaches to draw back the curtain. The screen goes black.

C3.6 The Ring Dur: 05.20

00.02 …[rain]

A rainy night. In the bedroom of a large old house, two teenage girls watch TV

00.32 …I don't care

Becca hands Katie the remote. Katie hits the off button

01.46 …it was…

Katie swallows

02.03 …phone rang

Becca's eyes widen

02.16 …just trying to scare me

Katie starts to choke, clutching at her throat

02.30 …[gurgling]

She grins!

02.40 …you guys do anything?

Katie looks blank

02.46 …you totally did, you're…

Becca whacks her with a cushion

02.52 *…[phone rings]*

Katie freezes. She looks at the clock

02.56 *…[brrrring]*

The hands point to 10

03.02 …[brrrring]

She stares at Becca

03.05 …there really is a tape?

Katie trembles. Barefoot, she pads downstairs. Becca follows.

03.17 …[brrrring]

The phone's in the hall

03.22…[brrrring]

Katie approaches nervously

03.34 *…residence*

Becca's smile fades. She turns sombrely to Katie and passes her the phone.

Katie's eyes are wide with fear

04.05 …we're about to go to sleep

She opens the fridge

04.35…yeah, yeah I will. [clonk]

She pours out juice

04.19…Bye Mom

In the lounge, the TV flicks on. Katie slowly turns towards it, face tense, mouth open a little, a frown creasing her brow. She steps through to the lounge. White static appears on screen

04.39…Becca, quit being a …

Katie looks around anxiously

04.44…where's the remote?

It's on the sofa. Katie picks it up and zaps off the TV

04.51 …huh!

She throws down the remote and turns to go…

04.55 …[static]

… stops and looks fearfully back, her breathing quick and shallow

05.13 …Becca quick!

She darts back, kneels by the TV and yanks out the plug. A shadow's reflected in the blank screen. She turns. Darkness

C4 Personal Details Sheet ID number Male/Female

Q 1. Would you describe yourself as having no/some/considerable useable vision?

Q 2. Are you registered: Blind Yes/No Partially Sighted Yes/No

If yes, at what age were you registered?

Q 3. Which of these best describes your sight with glasses or contact lenses if you normally use them?

A: I have no light perception

B: I can tell by the light where the windows are

C: I can see the shapes of furniture in the room

I can recognise a friend by sight alone if….

D: I'm close to their face

E: I'm at arms' length away

F: I'm on the other side of the room

G: I'm on the other side of the street.

Q 4. How long has your sight been this way?

Q5. Please state the medical name of your eye condition if known:

Q 6. Please state your age:

Q7. To what extent do you agree with the following statement on a scale of 1 – 5, where 1 = strongly disagree, 2= disagree, 3=neither agree nor disagree, 4= agree and 5=strongly agree? I am used to listening to Audio Description.

Q8. How many years of formal education have you had?
(up to GCSE/O-level = 12 years)

Thank You.

C5 Sample debriefing sheet (font size reduced for purposes of thesis)

Thank you for taking part in this session. The results will help me develop an understanding of the role AD plays in experiencing emotion. It forms part of my PhD research at Goldsmiths, University of London. It will also inform my work as an audio describer.

The main aims of this research are:

- To understand AD from the User's point of view
- To test whether AD can stimulate engagement with audiovisual media

If taking part has raised any issues for you or has left you with any questions or concerns please feel free to contact me, Louise Fryer, via Goldsmith's Psychology Dept. on 020 7919 7870, or by email: l.fryer@gold.ac.uk. Alternatively you may contact Dr. Jonathan Freeman who is supervising this research. He can also be contacted via Goldsmith's Psychology Dept. on 020 7919 7870. His email address is: j.freeman@gold.ac.uk The RNIB can offer emotional support and a listening ear via their helpline: 0303 123 9999

Thank you.

C6 Study 6 Correlations

For all tables: *. Correlation is significant at the 0.05 level (2-tailed).
**. Correlation is significant at the 0.01 level (2-tailed).

C6.1 Fear Clips

	Arousal	Sadness	Interest	Fear	Confusion
Sadness	-.340				
Interest	.517*	-.190			
Fear	.600**	-.060	.280		
Confusion	-.250	.040	-.110	.250	
Tension	.510*	-.230	.390	.486*	.020

Table C6.1 EES correlations: No AD

	Arousal	Sadness	Interest	Fear	Confusion
Sadness	-.380				
Interest	.040	.050			
Fear	.060	.010	.090		
Confusion	.360	.170	.050	.220	
Tension	.300	.010	.390	.270	.450

Table C6.2 EES correlations: TTS AD

	Arousal	Sadness	Interest	Fear	Confusion
Sadness	-.410				
Interest	.130	.010			
Fear	-.010	.050	-.170		
Confusion	-.310	.290	.030	.120	
Tension	.010	.060	.150	.170	.040

Table C6.3 EES correlations: HV AD

Fear clips continued:

For all tables: *. Correlation is significant at the 0.05 level (2-tailed).
**. Correlation is significant at the 0.01 level (2-tailed).

	Spatial	Engagement	Ecological	Fear
Engagement	.131			
Ecological	.459*	.490*		
Fear	.045	.658**	.209	
Tension	-.053	.618**	.332	.486*

Table C6.4 Target Emotion and Presence: No AD

	Spatial	Engagement	Ecological	Fear
Engagement	.281			
Ecological	.697**	.282		
Fear	.404	.363	.482*	
Tension	.460	.426	.352	.268

Table C6.5 Target Emotion and Presence: TTS AD

	Spatial	Engagement	Ecological	Fear
Engagement	.468*			
Ecological	.799**	.549*		
Fear	.108	.225	.121	
Tension	.223	.153	.130	.171

Table C6.6 Target Emotion and Presence: HV AD

Fear clips continued:

For all tables: *. Correlation is significant at the 0.05 level (2-tailed).
**. Correlation is significant at the 0.01 level (2-tailed).

	Fear	Confusion	Tension	HRV	Spatial	Engage	Ecol.
Confusion	.152						
Tension	.552*	-.067					
HRV	-.075	.050	.143				
Spatial	.083	-.181	-.134	.274			
Engage	.705**	-.389	.630**	.013	.131		
Ecological	.284	-.529*	.162	.231	.459*	.490*	
Negative	-.095	.331	.112	.197	-.362	-.084	-.244

Table C6.7 Target Emotions (Fear and Tension), Confusion, HRV and Presence: No AD

	Fear	Confusion	Tension	HRV	Spatial	Engage	Ecol.
Confusion	.215						
Tension	.504*	.188					
HRV	.096	.241	-.170				
Spatial	.496*	.199	.667**	.225			
Engage	.269	.339	.432	.029	.281		
Ecological	.528*	.268	.582*	.206	.697**	.282	
Negative	-.289	.075	-.109	.212	-.300	-.041	-.318

Table C6.8 Target Emotions (Fear and Tension), Confusion, HRV and Presence: TTS AD

Fear clips continued

	Fear	Confusion	Tension	HRV	Spatial	Engage	Ecol.
Confusion	.505*						
Tension	.688**	.217					
HRV	-.227	-.303	-.168				
Spatial	.141	-.280	.169	.215			
Engagement	.183	.002	.366	.197	.468*		
Ecological	.106	-.162	.155	.337	.799**	.549*	
Negative	-.042	.377	-.206	.372	-.152	-.061	-.023

*. Correlation is significant at the 0.05 level (2-tailed).
**. Correlation is significant at the 0.01 level (2-tailed).

Table C6.9 Target Emotions (Fear and Tension), Confusion, HRV and Presence: HV AD

C6. 2 Correlations: Sadness Clips

NB for all tables: *. Correlation is significant at the 0.05 level (2-tailed).**. Correlation is significant at the 0.01 level (2-tailed).

	Spatial	Engagement	Ecological	Negative	Arousal	Sadness	Interest
Engage	.505*						
Ecological	.560*	.790**					
Negative	-.135	.230	-.166				
Arousal	.321	.809**	.550*	.331			
Sadness	.504*	.644**	.693**	.147	.574*		
Interest	.489*	.807**	.639**	.113	.789**	.709**	
Confusion	-.161	.109	-.081	.418	.079	.019	.092

Table C6. 10 Correlations between Presence measures, target emotions (Arousal, Sadness, Interest) and Confusion: No AD

	Spatial	Engagement	Ecological	Negative	Arousal	Sadness	Interest
Engage	.876**						
Ecological	.811**	.760**					
Negative	-.384	-.274	-.344				
Arousal	.718**	.531*	.525*	-.323			
Sadness	.502*	.583**	.394	-.076	.286		
Interest	.833**	.720**	.691**	-.391	.795**	.598**	
Confusion	.137	.236	.201	-.336	.149	-.049	.221

Table C6. 11 Correlations between Presence measures, target emotions (Arousal, Sadness, Interest) and Confusion: TTS AD

Sadness clips continued

NB for all tables: *. Correlation is significant at the 0.05 level (2-tailed).**. Correlation is significant at the 0.01 level (2-tailed).

	Spatial	Engagement	Ecological	Negative	Arousal	Sadness	Interest
Engage	.720**						
Ecological	.830**	.735**					
Negative	-.290	-.334	-.454				
Arousal	.070	.243	-.021	.128			
Sadness	.121	.487*	.383	-.285	.239		
Interest	.201	.041	.215	.000	.434	.370	
Confusion	.288	.093	.254	.057	-.101	-.035	.194

Table C6. 12 Correlations between Presence measures, target emotions (Arousal, Sadness, Interest) and Confusion: HV AD

Appendix D

D1: Practical Implications for Audio Description: An Ecological Approach

The studies in this thesis have revealed as much about visual dominance of sighted people as about the multisensory experience of B&PS. In case of any doubt, one small anecdote on the subject of visual dominance may be of interest. One participant in Study 6 gradually lost her sight in her teens due to diabetic retinopathy. Gradually she accustomed herself to living with blindness. Some years later her diabetes necessitated receiving a kidney transplant. As kidney function returned, so did her sight. Experienced as she was in navigating using non-visual senses, she reported difficulty as her new (and poorly developed) sight tried to take over. This was despite her conscious efforts to use what, for her, had become more reliable sensory modalities. The problem for Audio Describers is that they, too, have the potential to overlook the contribution of non-visual modalities.

Elisabeth Weis (1984) describes a scene in Alfred Hitchcock's film *The Secret Agent* in which a dog's whine is included in the soundtrack, but no dog is visible in shot. The sound "represents" the murder of an innocent man. Weis argues that repeated experiments have shown that sighted audiences, and even film scholars, told to pay attention to the scene do not notice the dog's whine. However participants told specifically to listen to the soundtrack will always report on it, even though they have no knowledge of the plot. This is an indication of what she calls "the imbalance in our perceptual sensitivity, of our discriminating of visual information over the soundtrack" (p. 75).

American AD guidelines suggest that the first rule of description is to describe what you see: “w.y.s.i.w.y.s. – what you see is what you say” (Audio Description International, 2005). Yet if we were literally to describe everything in the visual array, users would be overwhelmed by detail. In their comparison of AD guidelines across countries, Rai et al. (2010, p. 71) suggest that the greatest challenge is how to choose “what not to describe”.

Redundancy has been a recurrent theme in terms of qualitative and quantitative findings of the studies presented here. Yet how does a describer know what is redundant? Gibson’s theory of affordances provides a useful guideline. As discussed in Chapter 2, Gibson argues that objects and environments are not neutral. They are not simply there for us to respond to as we will. Objects *afford* action. An object of a particular size, say a twig, affords us the opportunity to pick it up, perhaps to use as a tool, or as fuel for a fire. A larger object, the tree branch for example, may afford grasping but not carrying. An attached object, the tree itself, cannot be carried and nor, if the trunk is wide, can it be grasped. However, it may afford climbing. If it is a fruit tree, the fruit may afford eating. Gibson points out that animals are the most complex type of objects we perceive. He argues that an animal may be

> “…prey or predator, potential mate or rival, adult or young, one’s own young or another’s young. Moreover it may be temporarily asleep or awake, receptive or unreceptive, hungry or satiated. What the other animal affords is specified by its permanent features and its temporary

> state and it can afford eating or being eaten, copulation or fighting, nurturing or nurturance." (1979, p. 42)

Picking up such information is, according to Gibson, the purpose of sight. Sight enables us to move through an environment safely, engaging with that environment and the objects it contains. He suggests visual perception is crucial to that engagement. A surface that affords sitting must look as though it can be sat on. A glass floor, no matter how substantial, does not look like a supporting surface and the brain can be misled. Conversely, a glass door may look like an opening but does not afford passage, or a tilted room can confuse our ability to stand vertically (Asch et al., 1948).

Gibson explains that when we look at objects what we perceive are their affordances rather than their qualities. His list of qualities includes colour, texture, composition, size, shape, features of shape, mass, elasticity, rigidity and mobility. The list goes beyond what we might consider to be purely visual, and yet all can be ascertained at a glance. Texture, for example, is usually visible in that a coarser surface has a more speckled appearance; a perfectly smooth surface has a continuous colour. However, much of that visual information fails to register at a conscious level. Gibson suggests "we can discriminate the dimensions of difference if required to do so…but the special combination of qualities by which an object can be analysed is ordinarily not noticed" (1979, p. 134).

It is the role of the audio describer to notice what is ordinarily not noticed. Audio describers may pride themselves on their ability to do so. But how much of that detail should be included in a description? Gibson claims "However skilled an explicator one may become one will always, I believe, see more than one can say" (1979, p. 261). He gives the illustration of a cat on a mat. What we see is a cat obscuring some bits of the mat; the mat extending on each side of the cat, the mat supporting the cat, the floor that supports the mat and the cat, as it extends horizontally away from mat in each direction, the rigidity of the floor that is sufficient to take the weight of the mat and the cat, what the floor is made of and how smooth, flat or otherwise it may be. We also see where we are in relation to the cat and the mat. We even see parts of ourselves, the shadow of our nose, and perhaps our fringe or a strand of hair, our feet and hands and parts of our forearms. And we see whether the cat is sleeping or awake, whether it is friendly or twitching its tail. We see what colour the cat is, and perhaps some indication of its age and whether it has been in a fight or an accident and is missing part of its left ear… How does the concept of affordances help the describer choose what is relevant amongst all the visual information that we have at our disposal?

The question to ask is "what does the cat on the mat afford'? If the cat is twitching its tail, it is less likely to afford us the opportunity of stroking it. Similarly if its coat is scabby or crawling with fleas. If the cat features in a film we are describing, it may afford information about its owner – the Persian cat stroked by the Bond villain Blofeld in *To Russia with Love* (Saltzman et al., 1963), indicates Blofeld's interest in prestige and appearance. That the cat is long-haired, white and wears a jewel-encrusted collar tells us much about

Blofeld's circumstances, both financial and environmental. By contrast, in *Austin Powers: International Man of Mystery* (Myers et al., 1997) the hairless Sphynx cat adopted by the spoof villain, Dr. Evil, affords comparison with Blofeld in terms of power and over-weaning ambition, and humour in terms of abundant hair versus no hair. Although neither cat may affect the plot, the visual information we select can enhance the AD user's understanding of a key character and enrich their enjoyment of the film.

As noted in Chapter 1, affordances draw our attention, even if we are not consciously aware of it. In a film, the camera may linger on a knife – focusing our attention on what it affords and prompting the describer to include it in the AD. On a stage set, it will be left to the describer to consciously register the knife and mention it in their description, so the AD user is suitably primed. It is possible to include a great deal of visually-apprehended information as to the make, the colour, the age and condition of the knife. If time is short, however, simply stating that the knife is there may be sufficient, as AD users can infer what it affords. Dependent on the context it is likely to afford cooking or something more sinister.

However, the converse of this applies too. By mentioning something, the describer sets up an expectation of its importance. In detailing every object on a cluttered mantelpiece the AD may convey the impression that each will feature in the action: the candle in the candlestick will be lit; the clock will be wound; flowers will be placed in the vase; a photograph will be examined. An AD user may be waiting for the moment these props are brought into the limelight, and waste cognitive effort in doing so.

Surely the main problem for a describer is the danger of making a subjective selection from all the visual information at their disposal? This assumes a rationalist way of thinking about the world. The rationalist stance establishes the subjective view of the individual as distinct from the objective “facts” of the real world. Zahorik & Jenison (1998) point out that Gibson, together with the philosopher Heidegger, prefer to view existence as being inseparable from our interaction with the physical world around us. For the rationalist, perception is how we internalize the objective, external world. For Gibson and Heidegger, however, there is no objective/subjective distinction.

Heidegger (1962; 1977) argues that meaning is derived from interpretation that naturally arises from individual differences in the beliefs, language and experience of the interpreter. Furthermore, he proposes that we rarely have the opportunity for analytic (objective) reflection as we are constantly thrown into situations that require action. We do, rather than think. For Heidegger, like Gibson, objects are not “objective”. They are there to be used and may change their nature in the course of that manipulation. A sturdy stick may become a mallet for the duration it is used to drive a tent peg into the ground, but becomes a stick again when it is tossed onto the campfire. The external world of objects and environments is therefore inherently unstable.

The relevance of this to AD can be illustrated by the experience of describing a short film that showed a man in, what functionally at least, was a workshop. It was a small space, with shelves covering the walls. On the shelves were tools and other equipment. On one side there was a washing machine and a long workbench with a vice attached; at the far end was a sink, with a window looking out onto the garden. Beside the sink unit was a normal-sized backdoor, half-glazed. An external shot showed this space to be a garage attached to the main house. However, the term “garage” would have been misleading. It would introduce a schema in which a car was likely to be present, or at least a space in the centre large enough for a car to be parked. A garage is likely to have no windows and a wide “up-and-over” door. Although technically inaccurate, the term “workshop” summed up both the appearance and the function of this particular location much better.

Heidegger touches on a similar effect when he argues that a hammer used for hammering is “ready-to-hand” but if a “break down” occurs, for example it slips, then other aspects of the hammer become present-at-hand (e.g. we notice it is large and heavy and can injure the person who is hammering, or that the hammerhead is loose, and can fly off and break a window). This distinction between what an object is and how it is being used can be helpful in determining which parts of it we describe. The term “object” is used here in the Gibsonian sense, including animals and people. For one congenitally blind AD user then:

> "The fact of knowing that someone was 5'3" would be important but only in the sense that I would know or I would have an idea as to what his physical limitations would be."

And for another, when asked what aspects of a character he would like described:

> 'It's more facts that tell me about the person, like grey hair, balding, moustached. Things like that I've learned though literature and general life experience tell you something about the person."

When it comes to properties of objects, the distinction between quality and affordance is a useful one for describers, who are often too focused on what they see. As discussed in Chapter 2, Zahorik & Jenison use the example of a basketball, arguing that it is not best represented by the features "round", "orange" and "rubber" but rather by the fact it can be thrown, rolled or bounced. In AD, the colour of a ball may be less important than the type of ball. Knowing if the ball in question is a golf ball, a cricket ball, a tennis ball or a medicine ball allows the AD user to infer details of its size, its weight, how bouncy it is and to anticipate how it might be used.

Objects are features of our environment. Gibson suggests that our perception of the environment can be judged to be "true" if we are able to engage with it successfully. As evidenced in Chapter 7, sighted people process much of the critical detail implicitly. The describer's challenge comes in providing the right information *verbally* to people who cannot pick up that information for themselves. The describer has to consciously determine aspects of an object's affordance. For example, in describing a chair, a person without vision needs to know whether it has arms or not, or a low seat that may affect how far they need to bend, and whether the chair is unoccupied and therefore available for use. Chapter 7 in particular shed light on how someone with impaired vision engages with the environment. Such information enables a describer to provide those aspects of the visual array that the AD user will find useful or relevant.

A crucial element of the ecological approach is that the observer is not passive. There is a reciprocal relationship between the observer and what is observed. By extension, for those excluded from the process of direct observation owing to their vision loss, the describer affords information. But again, those in receipt of that information are not empty vessels waiting to be filled by the describer's words. The AD user will infer meaning based on his or her own experience. They may listen less keenly if the information provided is not useful or interesting or is too much or insufficient. They may turn off completely (mentally or by literally reaching for the off-switch on their TV remote or their headset) if they find the describer's voice distracting or hard to hear, especially as Study 6 suggests, if it is electronic speech. The describer, too, may be affected by knowledge of their audience. In instances of live AD, a description may be tailored to those attending the performance, taking into account their particular interests, sight characteristics or prior knowledge that might have been obtained from a pre-show workshop or touch tour.

While this lack of objectivity is generally regarded by the AD community as something to be avoided (see Rai et al., 2010) for Gibson it is an inevitable and necessary feature of existence. AD users confirm that, in the right way, what may be perceived as subjectivity on the part of the describer is not problematic. When asked about description of characters, one man put it this way:

> "Clothes and looks aren't so familiar to me as [being] someone blind from birth, but obviously clothes define the person, looks define the person – and I know this is a bit of a contentious issue - whether someone is fat or thin or Chinese-looking or black or short or lanky or pretty or ugly or horrible or lovely is I guess important to know, even though it may be subjective."

Gibson's approach is not without its limitations. In particular, he ascribes perception of affordances primarily to vision, taking little account of information picked up through non-visual modalities. Woods and Newell (2004), for example, point out that while vision can successfully identify a glass of water affording the opportunity to drink, it may need touch to confirm whether the water is too cold to drink comfortably. Similarly we can tell much about an object through sound: for example we can perceive its length (Carello et al., 1998; 2005), its size (Grassi, 2005) and whether or not it is within reach (Rosenblum et al., 1996). We can recognise natural sounds such as footsteps, applause, or a can being opened. We recognise voices, the spoken word and emotion in the voice of the speaker (Davis & Johnsrude, 2007). From this we can deduce much about the speaker including whether or not he or she is safe to approach.

The affordances of sound have been utilised to improve robotic navigation (Chu et al., 2009). Recognising affordances through non-visual modalities is important for the describer too, especially if they are to avoid supplying redundant information, for example describing the presence of a clock or a doorbell when the sound is self-evident (see Peli & Fine, 1996; Fryer, 2010b). By concentrating on affordances, it is also important not to forget the pure aesthetic enjoyment that visual qualities can provide. As one participant expressed it:

> "Because I've never experienced light in all its wonderful forms…I'm entirely fascinated … and I love to hear talk of what light looks like and colours of fireworks and I love it all…it doesn't bother me because I can't see it, I just love hearing about it."

An ecological approach to audio description helps clarify not only what should be described but why. It reminds us that visual perception provides information that is about more than simply the appearance of things; visual perception facilitates our engagement with objects in a particular environment. If the function of AD is to allow a blind or partially sighted user to do the same, for example interacting with props at a touch tour, or exhibits in a museum, then visual qualities that convey what those props or exhibits afford should be the ones that dominate the descriptive agenda. Similarly, in a virtual environment, affordances of characters, objects and locations are what blind and partially-sighted audiences need to know in order to follow the on-screen action.

Ultimately Gibson’s insights make it clear that we should not “say what we see” simply because we can see it.

References

Afonso, A., Blum, A., Katz, B. F., Tarroux, P., Borst, G., & Denis, M. (2010). Structural properties of spatial representations in blind people: Scanning images constructed from haptic exploration or from locomotion in a 3-D audio virtual environment. *Memory & Cognition*, *38*(5), 591-604.

Allen, G. L. (1999). Cognitive abilities in the service of wayfinding: A functional approach. *Professional Geographer 51*(4), 554–561.

Amedi, A., Floel, A., Knecht, S., Zohary, E., & Cohen, L. G. (2004). Transcranial magnetic stimulation of the occipital pole interferes with verbal processing in blind subjects. *Nature neuroscience*, *7*(11), 1266-1270

Appelhans, B. M., & Luecken, L. J. (2006). Heart rate variability as an index of regulated emotional responding. *Review of General Psychology, 10 (3),* 229 – 240.

Asch, S. E., & Witkin, H. A. (1948). Studies in space orientation. IV. Further experiments on perception of the upright with displaced visual fields. *Journal of Experimental Psychology, 38*, 455–477.

Atchley, R.A., Ilardi, S.S., Young, K.M., Stroupe, N.N., O'Hare, A.J., Bistricky, S.L., Collison, E., Gibson, L., Schuster, J., & Lepping, R.J. (2012). Depression reduces perceptual sensitivity for positive words and pictures. *Cognition & Emotion, 26* (8), 1359-1370.

Audio Description International. (2005). *Guidelines for audio description.* Retrieved from http://www.adinternational.org/ADIguidelines.html

Baños, R.M., Botella, C., Alcañiz, M., Liaño, V., Guerrero, B., & Rey, B. (2004). Immersion and Emotion: Their impact on the sense of presence. *CyberPsychology & Behavior, 7*(6): 734-741. doi:10.1089/cpb.2004.7.734.

Barfield, W., Zeltzer, D., Sheridan, T. B., & Slater, M. (1995). Presence and performance within virtual environments. In W. Barfield & T.A. Furness (Eds.) *Virtual environments and advanced interface design* (pp. 473–541). Oxford: Oxford University Press.

Barsalou, L. W. (2010). Grounded cognition: past, present, and future. *Topics in Cognitive Science*, *2* (4), 716-724.

Barsalou, L. W., Santos, A., Simmons, W. K., & Wilson, C. D. (2008). Language and simulation in conceptual processing. *Symbols, embodiment, and meaning*, 245-283.

Beer, A. L., Plank, T., & Greenlee, M. W. (2011). Diffusion tensor imaging shows white matter tracts between human auditory and visual cortex. *Experimental Brain Research*, *213*(2-3), 299-308.

Bedny, M., Pascual-Leone, A., Dodell-Feder, D., Fedorenko, E., & Saxe, R. (2011). Language processing in the occipital cortex of congenitally blind adults. *Proceedings of the National Academy of Science, U.S.A.*, *108*(11), 4429-34.

Belin, P., Zatorre, R.J., Lafaille, P., Ahad, P., & Pike, B. (2000). Voice-selective areas in human auditory cortex. *Nature, 403*, 309 –312.

Bennett, A. T. D. (1996). Assessing models of spatial memory: the cognitive map. *Journal of Experimental Biology,199*, 219–224.

Bertelson, P., & de Gelder, B. (2004). The psychology of multi-modal perception. In J. Driver & C. Spence (Eds.) *Semantic Encoding in working memory: Cross modal space and cross modal attention.* London: Oxford University Press.

Betts, G. H. (1909). *The distribution and functions of mental imagery (No. 26).* Teachers College, Columbia University Press.

Bhattacharya, J. (2009). Increase of universality in human brain during mental imagery from visual perception. *PloS one*, *4*(1), e4121

Bidet-Caulet, A., Voisin, J., Bertrand, O., & Fonlupt, P. (2005). Listening to a walking human activates the temporal biological motion area. *Neuroimage*, *28*(1), 132-139.

Bigelow, A. E. (1991). Spatial Mapping of familiar locations in blind children. *Journal of Visual Impairment and Blindness 85*, 113-17.

Bigelow, A. E. (1996). Blind and sighted children's spatial knowledge of their home environments. *International Journal of Behavioural Development, 19*, 797-816.

Biocca, F. (1997). The cyborg's dilemma: embodiment in virtual environments. *Journal of Computer-Mediated Communication 3* (2). Retrieved from http://jcmc.indiana.edu/vol3/issue2/biocca2.html

Biocca, F. (2003, May). *Can we resolve the book, the physical reality, and the dream state problems? From the two-pole to a three-pole model of shifts in presence.* Paper presented at the EU Future and Emerging Technologies Presence Initiative Meeting, Venice. Retrieved from http://www.mindlab.org/images/d/DOC705.pdf

Blasch, B.B., Wiener, W.R., & Welsh, R.L. (1997). *Foundations of orientation and mobility* (2nd ed.). New York, NY: AFB Press

Bolls, P. D. (2002). I can hear you but can I see you? The use of visual cognition during exposure to high imagery radio advertisements. *Communication Research*, 29, 537-563.

Bonneel, N., Suied, C., Viaud-Delmon, I., & Drettakis, G. (2009). Bimodal perception of audio-visual material properties for virtual environments. *ACM Transactions on Applied Perception* (Accepted with minor revisions).

Botvinick, M., & Cohen, J. (1998). Rubber hands "feel" touch that eyes see. *Nature*, *391*(6669), 756-756.

Bracken, C. C. (2005). Presence and image quality: The case of high definition television. *Media Psychology*, *7*, 191– 205.

Bracken, C. C. (2010). Sounding out small screens and telepresence: The impact of audio, screen size, and pace. *Journal of Media Psychology, 22* (3), 125 -137

Brambring, M. (1982). Language and geographic orientation for the blind. In H.J. Jarvella and W. Klein (Eds.) *Speech, Place and Action: Studies in Deixis and Related Topics* (pp. 161-182). New York: Wiley.

Bransford, J. D., & Johnson, M. K. (2004). Contextual prerequisites for understanding: Some investigations of comprehension and recall. In D. A. Balota, & E. K. Marsh (Eds.) *Cognitive psychology: Key readings* (pp. 431-439). New York, NY, US: Psychology Press.

Bremner, A. J., Caparos, S., Davidoff, J., de Fockert, J., Linnell, K.J., & Spence, C. (2012). "Bouba" and "Kiki" in Namibia? A remote culture make similar sound-shape matches, but different taste-shape matches to Westerners. *Cognition*, http://dx.doi.org/10.1016/j.cognition.2012.09.007

Bruner, J. S., & Postman, L. (1949). On the perception of incongruity: a paradigm. *Journal of Personality,18,* 206–223.

Bulu, S., T. (2012). Place presence, social presence, co-presence, and satisfaction in virtual worlds. *Computers & Education, 58*, 154–161.

Burton, H., Snyder, A. Z., Conturo, T. E., Akbudak, E., Ollinger, J. M., & Raichle, M. E. (2002a). Adaptive changes in early and late blind: a fMRI study of Braille reading. *Journal of Neurophysiology*, *87*(1), 589-607.

Burton, H., Snyder, A. Z., Diamond, J. B., & Raichle, M. E. (2002b). Adaptive changes in early and late blind: a FMRI study of verb generation to heard nouns. *Journal of Neurophysiology, 88*(6), 3359-3371.

Buxton, W. (1992). Telepresence: Integrating shared task and person spaces. *Proceedings of Graphics Interface*, 123-129.

Canning, C., & Fryer, L. (2010, October). *Making sense of touch*. Paper presented at the In Touch With Art conference. St. Dunstans. Victoria and Albert Museum, London.

Carello, C., Anderson, K. L., & Kunkler-Peck, A. J. (1998). Perception of object length by sound. *Psychological Science 9 (3).* 211 – 214.

Carello, C., Wagman, J. B., & Turvey, M. T. (2005). Acoustic specification of object properties. In J. D. Anderson & B. Fisher Anderson (Eds.) *Moving Image Theory: Ecological Considerations* (pp.79 – 104). Carbondale, IL: Southern Illinois University Press.

Carroll, J.B. (1993). *Human cognitive abilities: A survey of factor analytic studies.* New York: Cambridge University Press.

Cattaneo, Z., Fantino, M., Silvanto, J., Tinti, C., Pascual-Leone, A., & Vecchi, T. (2010). Symmetry perception in the blind. *Acta Psychologica, (3*), 398 - 402.

Cattaneo, Z., & Vecchi, T. (2011). *Blind vision: the neuroscience of visual impairment*. Cambridge Massachussets: MIT Press.

Cattaneo, Z., Vecchi, T., Cornoldi, C., Mammarella, I., Bonino, D., Ricciardi, E., et al. (2008). Imagery and spatial processes in blindness and visual impairment. *Neuroscience and Biobehavioral Reviews, 32*(8), 1346–1360.

Chandler, D.W., Grantham, D.W., & Leek, M.R. (1993). Auditory spatial resolution in the horizontal plane as a function of reference angle: Microstructure of the azimuth function. *Journal of the Acoustical Society of America, 93*, 2350 (A).

Chen, C. (2000). Individual differences in a spatial-semantic virtual environment. *Journal of the American Society for Information Science*, 51, 529–542.

Cherry, E. C. (1953). Some experiments on the recognition of speech with one and two ears. *Journal of the Acoustical Society of America, 25,* 975–979

Chu, S., Narayanan, S., & Kuo, C.-C. J., 2009. Environmental sound recognition with time-frequency audio features. *IFEE transactions on audio, speech and language processing, 17 (6)*, 1142 – 1158.

Collignon, O., Renier, L., Bruyer, R., Tranduy, D., & Veraart, C. (2006). Improved selective and divided spatial attention in early blind subjects. *Brain Research*, *1075*(1), 175-182.

Collignon, O., Vandewalle, G., Voss, P., Albouy, G., Charbonneau, G., Lassonde, M., & Lepore, F. (2011). Functional specialization for auditory–spatial processing in the occipital cortex of congenitally blind humans. In *Proceedings of the National Academy of Sciences, 108* (11), 4435 - 4440.

Collignon, O., Voss, P., Lassonde, M., & Lepore, F. (2009). Cross-modal plasticity for the spatial processing of sounds in visually deprived subjects. *Experimental brain research, 192*(3), 343-358.

Conway, M. A. (2001). Sensory–perceptual episodic memory and its context: Autobiographical memory. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences*, *356* (1413), 1375-1384.

Crisinel, A. S., Jones, S., & Spence, C. (2012). The sweet taste of maluma: Crossmodal associations between tastes and words. *Chemosensory Perception, 5*, 266 – 273.

Crook, T. (1999). *Radio drama theory and practice*. London: Routledge.

Crossman, A. R., & Neary, D. (2010). *Neuroanotomy: An illustrated colour text*. 4th edition. Edinburgh. London: Churchill Livingstone.

Cryer, H., & Home, S. (2009). *User attitudes towards synthetic speech for talking books*. In Report no 7, RNIB Centre for Accessible Information: Birmingham.

Csikszentmihalyi, M. (1991). *Flow: The psychology of optimal experience.* New York: Harper & Row.

Dale, N. (2013, May). *Infants with visual impairment and their parents – how they engage with each other*. Mary Kitzinger Trust workshop. Lecture conducted from UCL Institute of Child Health, London.

Davis, M. H., & Johnsrude, I. S. (2007). Hearing speech sounds: Top-down influences on the interface between audition and speech perception. *Hearing Research,* 229, 132 – 147.

Davis, R. (1961). The fitness of names to drawings: a cross-cultural study in Tanganyika. *British Journal of Psychology, 52*, 259-267.

de Haan, E. H., & Cowey, A. (2011). On the usefulness of 'what' and 'where' pathways in vision. *Trends in cognitive sciences*, *15*(10), 460-466.

de Rijk, A. E., Schreurs, K. M., & Bensing, J. M. (1999). Complaints of fatigue: Related to too much as well as too little external stimulation? *Journal of Behavioral Medicine, 22*, 549–573.

Delić, V., & Sedlar, N. V. (2010). Stereo presentation and binaural localization in a memory game for the visually impaired. In *Lecture Notes in Computer Sciences, LNCS 5967* (pp. 354 – 363). Heidelberg: Springer.

Deroy, A.-S., & Valentin, D. (2011). Tasting liquid shapes: Investigating the sensory basis of cross-modal correspondences. *Chemosensory Perception, 4*, 80 – 90.

Dillon, C. (2006). *Emotional responses to immersive media*. (Unpublished doctoral dissertation). Goldsmiths College, University of London, UK.

Dillon C., Keogh, E., Freeman, J., & Davidoff, J. (2001). *Presence: Is your heart in it?* Paper presented at the Fourth Annual International Workshop on Presence. Philadelphia, PA, USA. Retrieved from http://nimbus.temple.edu/~mlombard/P2001/dillon.htm

Douglas, G., Corcoran, C., & Pavey, S. (2006). *Network 1000: Opinions and circumstances of visually impaired people in Great Britain: report based on over 1000 interviews.* Visual Impairment Centre for Teaching and Research, University of Birmingham.

Draper, J. V., Kaber, D. B., & Usher, J. M. (1998). Telepresence. *Human Factors, 40* (30), 354 – 375.

Driver, J., & Spence, C. (2004). Crossmodal spatial attention: Evidence from human performance. In C. Spence & J. Driver (Eds.), *Crossmodal space and crossmodal attention* (pp. 179–220). Oxford, UK: Oxford University Press.

Eardley, A.F., & Pring, L. (2006). Remembering the past and imagining the future: a role for nonvisual imagery in the everyday cognition of blind and sighted people. *Memory 14*, 925 – 936

Eardley, A. F., & Pring, L. (2011). Exploring the impact of sucking sweets on flavour imagery. *Journal of Cognitive Psychology, 23* (7), 811 – 817.

Evans, E.J., & Pearson, R. (2009). Boxed out: visually impaired audiences, audio description and the cultural value of the television image. *Participations: Journal of Audience and Reception Studies. 6*(2), 373 – 402.

Fauconnier, G. (1999). Methods and generalizations. *Cognitive linguistics: foundations, scope and methodology* (pp. 95-128). Berlin: Mouton de Gruyter.

Feist, M. I., & Gentner, D. (2007). Spatial language influences memory for spatial scenes. *Memory & Cognition*, *35*(2), 283-296.

Fendrich, R., Wessinger, C. M., & Gazzaniga, M. S. (1992). Residual vision in a scotoma: implications for blindsight. *Science*, *258*(5087), 1489-1491.

Fiehler, K., & Rösler, F. (2010). Plasticity of multisensory dorsal stream functions: evidence from congenitally blind and sighted adults. *Restorative Neurology and Neuroscience*, *28*(2), 193-205.

Foo, P. Warren, W.H., Duchon, A., & Tarr, M. J. (2005). Do humans integrate routes into a cognitive map? Map- versus landmark-based navigation of novel shortcuts. *Journal of Experimental Psychology: Learning, memory and cognition. 31* (2), 195-215.

Foulke, E. (1971). The perceptual basis for mobility. *Research Bulletin, American Foundation for the Blind, 23*, 1-8.

Freeman, J. (2004). Implications for the measurement of presence from convergent evidence on the structure of presence. In *Information Systems Division* at the annual meeting of the International Communication Association, New Orleans, LA.

Freeman, J., & Avons, S. E. (2000, June). Focus group exploration of presence through advanced broadcast services. In *Electronic Imaging* (pp. 530-539). International Society for Optics and Photonics.

Freeman, J., Avons, S. E., Pearson, D. E., & IJsselsteijn, W. A. (1999) Effects of sensory information and prior experience on direct subjective ratings of presence. *Presence: Teleoperators and Virtual Environments*, *8*, 1 - 13.

Freeman, J., Lessiter, J., Pugh, K., & Keogh, E. (2005). When Presence and emotion are related, and when they are not. In *8th Annual International Workshop on Presence, September* (pp. 21-23).

Fryer, L. (2009). *Talking Dance: the audio describer's guide to dance in theatre*. ADA publications, Apt Description Series (3). ISBN 978-0-9560306-2-7.

Fryer, L. (2010a, October). *Being there: Audio describing live events*. Paper presented at the 9th Languages and the Media Conference, Berlin.

Fryer, L. (2010b). Audio description as audio drama: a practitioner's point of view. *Perspectives: Studies in Translatology*, 18(3). 205-213.

Fryer, L., & Freeman, J. (2012a). Presence in those with and without sight: implications for virtual reality and audio description. *Journal of CyberTherapy & Rehabilitation. 5*, 15-23.

Fryer, L., & Freeman, J. (2012b). Cinematic language and the description of film: Keeping AD users in the frame. *Perspectives: Studies in Translatology*, DOI:10.1080/0907676X.2012.693108

Gallace, A., Boschin, E., & Spence, C. (2012). On the taste of "bouba" and "kiki" An exploration of word-food associations in neurologically normal participants. *Cognitive Neuroscience*, *2* (1), 34-46

Galton, F. (1880). I.- STATISTICS OF MENTAL IMAGERY. *Mind*, (19), 301-318.

Gerber, E. (2007). Seeing isn't believing: Blindness, race and cultural literacy. *Senses and Society 2* (1), 27 - 40.

Gerrig, R. J. (1993). *Experiencing Narrative Worlds: On the Psychological Activities of Reading*. New Haven: Westview Press.

Gibson, J.J. (1979). *The Ecological Approach to Visual Perception*. Boston: Houghton Mifflin (3rd edition. 1986)

Gissurarson, L. R. (1992). Reported auditory imagery and its relationship with visual imagery. *Journal of Mental Imagery, 16*, 117 - 122.

Glenberg, A. M. (1997). What is memory for? *Behavioral and Brain Sciences, 20*, 1-55.

Goddard, L., Pring, L., & Felmingham, N. (2005). The effects of cue modality on the quality of personal memories retrieved. *Memory 13* (1), 79 - 86.

Goldreich, D., & Kanics, I. M. (2003). Tactile acuity is enhanced in blindness. *Journal of Neuroscience*, *23*(8), 3439-3445.

Goldreich, D., & Kanics, I. M. (2006). Performance of blind and sighted humans on a tactile grating detection task. *Perception & Psychophysics*, *68*(8),1363-1371.

Goldstein, J., Davidoff, J., & Roberson, D. (2009). Knowing color terms enhances recognition: Further evidence from English and Himba. *Journal of Experimental Child Psychology 102*, 219 – 238.

Golledge, R. G. (1993). Geography and the disabled - a survey with special reference to vision impaired and blind populations. *Transactions of the Institute of British Geographers*, 18(1), 63-85.

Golledge, R. G., Marston, J., Loomis, J., & Klatzky, R . L. (2004). Stated preferences for components of a personal guidance system for non-visual navigation. *Journal of Visual Impairment and Blindness*, *98,* 135–47.

Goodale, M. A., & Milner, A. D. (1992). Separate visual pathways for perception and action. *Trends in Neurosciences*, *15*(1), 20-25.

Gougoux, F., Lepore, F., Lassonde, M., Voss, P., Zatorre, R.J., & Belin, P. (2004). Pitch discrimination in the early blind. *Nature*, *430*, 309-1309.

Grassi, M. (2005). Do we hear size or sound? Balls dropped on plates. *Perception and Psychophysics 67(2),* 274-284

Green, M.C., & Brock, T.C. (2002). In the mind's eye: Transportation-imagery model of narrative persuasion. In M.C. Green, J.J. Strange, & T.C. Brock (Eds.), *Narrative impact. Social and Cognitive Foundations* (pp. 315–341). Mahwah, NJ: Erlbaum.

Green, M.C., Brock, T.C., & Kaufman, G.F. (2004). Understanding media enjoyment: The role of transportation into narrative worlds. *Communication Theory, 14*, 311–327.

Gross, J. & Levenson, R. W. (1995). Emotion elicitation using films. *Cognition and Emotion, 9,* 87 – 108.

Gysbers, A., Klimmt, C., Hartmann, T., Nosper, A., & Vorderer, P. (2004). Exploring the book problem: Text design, mental representations of space, and spatial presence in readers. In M. A. Raya & B. R. Solaz (Eds.), *Seventh Annual International Workshop: Presence 2004.* Universidad Politecnica de Valencia.

Hagemann, D., Waldstein, S. R., & Thayer, J. F. (2003). Central and autonomic nervous system integration in emotion. *Brain and Cognition*, *52*(1), 79-87.

Hartcher-O'Brien, J., Levitan, C., & Spence, C. (2010). Extending visual dominance over touch for input off the body. *Brain Research*, *1362*, 48-55.

Hecht, D., & Reiner, M. (2007). Field dependency and the sense of object-presence in haptic virtual environments. *CyberPsychology and Behavior, 10* (2), 243–251.

Heeter, C. (1992). Being there: The subjective experience of presence. *Presence*, 1(2), 262–271.

Heidegger, M. (1962). *Being and time*. (J. Macquarrie & E. Robinson, Trans.). San Francisco: Harper Collins. (Original work published 1927).

Heidegger, M. (1977). *The question concerning terminology*. (W. Lovitt, Trans.). New York: Garland Publishing. (Original work published 1952).

Heller, M. A., Wilson, K., Steffen, H., Yoneyama, K., & Brackett, D. D. (2003). Superior haptic perceptual selectivity in late-blind and very-low-vision subjects. *Perception*, *32*(4), 499-512.

Henderson, J. M., Malcolm, G. L., & Schandle, C. (2009). Searching in the dark: Cognitive relevance drives attention in real-world scenes. *Psychonomic Bulletin & Review, 16* (5), 850 – 856.

Hill, E. W., & Ponder, P. (1976). *Orientation and mobility techniques*. New York: American Foundation for the Blind.

Hobson, R. P. (1993). *Autism and the development of mind*. Hillsdale, NJ: Lawrence Erlbaum Associates.

Hobson, P. R., & Bishop, M. (2003). The pathogenesis of autism: insights from congenital blindness. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences, 358*(1430), 335-344.

Holtby, R. C., & d'Angiulli, A. (2012). The effects of interference on recognition of haptic pictures in blindfolded sighted participants: The modality of representation of haptic information. *Scandinavian Journal of Psychology, 53* (2), 112–118.

Hollins, M. (1985). Styles of mental imagery in blind adults. *Neuropsychologia, 23*(4), 561 - 566.

Hötting, K., Rösler, F., & Röder, B. (2004a). Altered auditory-tactile interactions in congenitally blind humans: An event-related potential study. *Experimental Brain Research. 159*, 1432-1106.

Hötting, K., & Röder, B. (2004b). Hearing cheats touch, but less in congenitally blind than in sighted individuals. *Psychological Science*, *15*(1), 60-64.

Hubbard, T. L. (2010) Auditory imagery: Empirical findings. *Psychological Bulletin 136* (20), 302-309.

Hüfner, K., Stephan, T., Flanagin, V. L., Deutschländer, A., Stein, A., Kalla, R., ... & Brandt, T. (2009). Differential effects of eyes open or closed in darkness on brain activation patterns in blind subjects. *Neuroscience letters, 466*(1), 30-34.

Hull, J. M. (2001) *On sight and insight, a journey into the world of blindness*. Oxford, Oneworld.

Huttenlocher, P. R. (2009). *Neural plasticity: The effects of environment on the development of the cerebral cortex*. Harvard University Press.

Igareda, P., & Maiche, A. (2009). Audio Description of emotions in films using eye tracking. In *Proceedings of the Symposium on Mental States, Emotions and their Embodiment* (pp. 20 - 23).

IJsselsteijn, W. A. (2002). Elements of a multi-level theory of presence: Phenomenology, mental processing and neural correlates. In *Proceedings of Presence 2002* (pp. 245 - 259).

IJsselsteijn, W. A., de Kort, Y. A. W., & Haans, A. (2006). Is this my hand I see before me? The rubber hand illusion in reality, virtual reality, and mixed reality. *Presence: Teleoperators and Virtual Environments*, *15*(4), 455 - 464.

IJsselsteijn, W. A., de Ridder, H., Freeman, J., & Avons, S. (2000). Presence: Concept, determinants and measurement. *Proceedings of the SPIE, Human*

Vision and Electronic Imaging V, 3959, 520 - 529.

IJsselsteijn, W. A., Freeman, J., & de Ridder, H. (2001). Presence: Where are we? *CyberPsychology and Behavior, 4*, 179 - 182.

Jacobson, D. (2001). Presence revisited: Imagination, competence, and activity in text-based virtual worlds. *CyberPsychology & Behavior*, *4*(6), 653 - 673.

Johnson, M. K., Hashtroudi, S., & Lindsay, D. S. (1993). Source monitoring. *Psychological Bulletin*, *114*(1), 3.

Jones, M. (2007). Presence as external versus internal experience: How form, user, style, and content factors produce presence from the inside. *Proceedings of the Tenth Annual International Meeting of the Presence Workshop, Barcelona, Spain* (pp. 115– 126). Retrieved from http://www.temple.edu/ispr/prev_conferences/proceedings/2007/Jones.pdf

Jurnet, I.A., Beciu, C.C., & Maldonado, J.G. (2005). Individual differences in the sense of presence. *Presence:Teleoperators, and Virtual Environments,* 133-142.

Kelso, M. T., Weyhrauch, P., & Bates, J. (1993). Dramatic Presence. *Presence: Teleoperators and Virtual Environments, 2*(1), 1–15.

Kim, K. H., Bang, S. W., & Kim, S. R. (2004). Emotion recognition system using short-term monitoring of physiological signals. *Medical and Biological Engineering and Computing*, *42*(3), 419-427.

King, A. J. (2004). Development of multisensory spatial integration. In C. Spence & J. Driver (Eds.), *Crossmodal space & crossmodal attention* (pp. 1-24). Oxford, UK: Oxford University Press.

Kitchin, R. & Blades, M. (2002). *The cognition of geographic space*. I.B. Tauris. London. New York.

Klatzky, R.L., Marston, J.R., Giudice, N.A., Golledge, R.G., & Loomis, J. M. (2006). Cognitive load of navigating without vision when guided by virtual sound versus spatial language. *Journal of Experimental Psychology: Applied, 17* (4), 223-232.

Klimmt, C., & Vorderer, P. (2003). Media Psychology “is not yet there”: Introducing theories on media entertainment to the Presence debate. *Presence: Teleoperators and Virtual Environments, 12* (4), 346–359.

Koch, C. & Tsuchiya, N. (2006). Attention and consciousness: two distinct brain processes. *Trends in Cognitive Sciences 11* (1), 16 – 22.

Köhler, W. (1947). *Gestalt psychology* (2nd edn.). New York: Liveright Publishing.

Koppen, C., & Spence, C. (2007) Assessing the role of stimulus probability on the Colavita visual dominance effect. *Neuroscience Letters 418*, 266 – 271.

Koriat, A., & Levy, I. (1979). Figural symbolism in Chinese ideographs. *Journal of Psycholinguistic Research, 6*, 93-103.

Kraft, R. N. (1986). The role of cutting in the evaluation and retention of film. *Journal of Experimental Psychology: Learning, Memory and Cognition, 12* (1), 155-163.

Kraut, R. E., Gergle, D., & Fussell, S. R. (2002). The use of visual information in shared visual spaces: Informing the development of virtual co-presence. In *Proceedings of the 2002 ACM conference on Computer supported cooperative work,* 31-40.

Kreifelts, B., Ethofer, T., Grodd, W., Erb, M., & Wildgruber, D. (2007). Audiovisual integration of emotional signals in voice and face: an event-related fMRI study. *Neuroimage*, *37*(4), 1445-1456.

Kreifelts, B., Wildgruber, D., & Ethofer, T. (2013). Audiovisual integration of emotional information from voice and face. In *Integrating Face and Voice in Person Perception* (pp. 225-251). New York: Springer.

Krejtz, I., Sarkowska, A., Krejtz, K., Walczak, A., & Duchowski, A. (2012). Audio description as an aural guide of children's visual attention: evidence from an eye-tracking study. In *Proceedings of the Symposium on Eye Tracking Research and Applications*, 99 – 106, New York.

Kuhn, T. S. (1970). *The structure of scientific revolutions*. Chicago: Chicago University Press.

Kupers, R., & Ptito, M. (2013). Compensatory plasticity and cross-modal reorganization following early visual deprivation. *Neuroscience and Biobehavioral Reviews* http://dx.doi.org/10.1016/j.neubiorev.2013.08.001

Lacey, S., Tal, N., Amedi, A., & Sathian, K. (2009). A putative model of multisensory object representation. *Brain topography, 21*(3-4), 269-274.

Laurienti, P. J., Burdette, J. H., Wallace, M. T., Yen, Y. F., Field, A. S., & Stein, B. E. (2002). Deactivation of sensory-specific cortex by cross-modal stimuli. *Journal of Cognitive Neuroscience, 14(3),* 420-429.

Lavie, N., Hirst, A., de Fockert, J. W., & Viding, E. (2004). Load theory of selective attention and cognitive control. *Journal of Experimental Psychology: General 133*(3), 339–354. doi: 10.1037/0096-3445.133.3.339.

Lenci, A., Baroni, M., Cazzolli, G., & Marotta, G. (2013). BLIND: a set of semantic feature norms from the congenitally blind. *Behavior Research Methods*, 1-16.

Lessiter, J., Freeman, J., Keogh, E. & Davidoff, J. (2001). A cross-media presence questionnaire: The ITC Sense of Presence Inventory. *Presence: Teleoperators, and Virtual Environments, 10*(3), 282-297.

Levenson, R. W. (2003). Blood, sweat, and fears: The autonomic architecture of emotion. *Annals of the New York Academy of Sciences, 1000*, 348-366.

Lewald, J. (2002). Opposing effects of head position on sound localization in blind and sighted human subjects. *European Journal of Neuroscience*, *15*(7), 1219-1224.

Linder, G., & Martinez, J. (2011, March). *Audio description with a TTS application*. Paper presented at the Advanced Research Seminar on Audio Description, Barcelona.

Lombard, M., & Ditton, T. (1997). At the heart of it all: the concept of presence. *Journal of Computer Communication*, 3 (2), 1-39.

Lopez, J., & Pauletto, S. (2009). The design of an audio film: portraying story, action and interaction through sound. *Journal of Music and Meaning, (8).* Accessed online: http://www.musicandmeaning.net/issues/showArticle.php?artID=8.2

Maiworm, M., Bellantoni, M., Spence, C., & Röder, B. (2012). When emotional valence modulates audiovisual integration. *Attention, Perception, & Psychophysics*, *74*(6), 1302-1311.

Mandryk, R. L., & Atkins, M. S. (2007). A fuzzy physiological approach for continuously modeling emotion during interaction with play technologies. *International Journal of Human-Computer Studies, 65(4)*, 329-347.

Mandryk, R. L., Inkpen, K. M., & Clavert, T. W. (2005). Using psychophysiological techniques to measure user experience with entertainment technologies. *Journal of Behaviour and Information Technology 25(2)*, 141-158

Mannan, S. K., Ruddock, K. H., & Wooding, D. S. (1997). Fixation sequences made during visual examination of briefly presented 2D images. *Spatial Vision, 11*(2), 157-178.

Mantovani, G., & Riva, G. (1999). "Real" presence: How different ontologies generate different criteria for presence, telepresence, and virtual presence. *Presence, 8* (5), 540–550.

Marcell, M., Malatanos, M., Leahy, C., & Comeaux, C. (2007). Identifying, rating and remembering environmental sound events. *Behavioural Research Methods, 39* (30) 561-569.

Marks, D. F. (1999). Consciousness, mental imagery and action. *British Journal of Psychology, 90*, 566-585.

Martin, A. (2007). The representation of object concepts in the brain. *Annual Review of Psychology, 58*, 25–45.

Martinsen, H., Tellevik, J.M., Elmerskog, B., & Storlilokken, M. (2007). Mental effort in mobility route learning. *Journal of Visual Impairment & Blindness, 101*, 327–338.

Maurer, D., Pathman,T., & Mondlock, C.J. (2006). The shape of boubas: sound-shape correspondences in toddlers and adults. *Developmental Science 9* (3), 316-322.

McDonald, J. J., Teder-Sälejärvi, W. A., & Hillyard, S. A. (2000). Involuntary orienting to sound improves visual perception. *Nature, 407*(6806), 906-908.

McGreevy, M. W. (1994). *An ethnographic object-oriented analysis of explorer presence in a volcanic terrain environment: Claims and evidence*. National Aeronautics and Space Administration, Ames Research Center.

McKelvie, S. J. (1995). Emotional expression in upside‐down faces: Evidence for configurational and componential processing. *British Journal of Social Psychology*, *34*(3), 325-334.

McMahan, A. (2003). Immersion, engagement, and presence: A method for analyzing 3-D video games. In M. J. P. Wolf and B. Perron (Eds.), *The Video Game Theory Reader* (pp. 67 – 86). New York, Routledge, 67-86

Merabet L. B., & Pascual-Leone A. (2009). Neural reorganization following sensory loss: The opportunity of change. *Nature Reviews Neuroscience, 11 (1)*, 44–52.

Miller, D. W., & Marks, L. J. (1992). Mental imagery and sound effects in radio commercials. *Journal of Advertising, 21(4),* 83-93.

Miller, G. A. (1956). The magical number seven, plus or minus two: some limits on our capacity for processing information. *Psychological review, 63*(2), 81.

Minsky, M. (1980). Telepresence. *Omni*, 2, 45–51.

Minshull, D. (Producer). (2011, February 28). *Radio 3 The Essay: Listener, they wore it*. Tracy Chevalier. London: BBC. Retrieved from http://www.bbc.co.uk/programmes/b00xnbjw

Minter, M., Hobson, R. P., & Bishop, M. (1998). Congenital visual impairment and 'theory of mind'. *British Journal of Developmental Psychology*, *16*(2), 183-196.

Mishkin, M., Ungerleider, L. G., & Macko, K. A. (1983). Object vision and spatial vision: two cortical pathways. *Trends in Neurosciences*, *6*, 414-417.

Morris, C. D., Bransford, J. D., & Franks, J. J. (1977). Levels of processing versus transfer appropriate processing. *Journal of Verbal Learning and Verbal Behavior*, *16*(5), 519-533.

Mousavi, S. Y., Low, R., & Sweller, J. (1995). Reducing cognitive load by mixing auditory and visual presentation modes. *Journal of Educational Psychology, 87*(2), 319.

Murphy, S., van Velsen, J., & de Fockert, J. W. (2012) The role of perceptual load in action affordance by ignored objects. *Psychonomic Bulletin & Review 19* (6), 1122-1127

Murray, J. (1997). *Hamlet on the Holodeck: The Future of Narrative in Cyberspace.* Cambridge, MA: The MIT Press.

Murray, M. M., Molholm, S., Michel, C. M., Heslenfeld, D. J., Ritter, W., Javitt, D. C., ... & Foxe, J. J. (2005). Grabbing your ear: rapid auditory–somatosensory multisensory interactions in low-level sensory cortices are not constrained by stimulus alignment. *Cerebral Cortex*, *15*(7), 963-974.

National Health Service. (2011). *Registered blind and partially sighted people, England, year ending 31 March 2011*. Retrieved from http://www.ic.nhs.uk/pubs/blindpartiallysighted11

Neisser, U. (1976). *Cognition and reality. Principles and implication of cognitive psychology.* San Francisco: Freeman.

Neuville, M. I., & Trassoudaine, L. (2009). A wayfinding pilot study: the use of the Intelligent Public Vehicle by people with visual impairment. *British Journal of Visual Impairment 27*, 65 – 74.

Ng, H. K., Kalyuga, S., & Sweller, J. (2013). Reducing transience during animation: a cognitive load perspective. *Educational Psychology,* (ahead-of-print), 1-18.

Ngo, M. K., Misra, R., & Spence, C. (2011). Assessing the shapes and speech sounds that people associate with chocolate samples varying in cocoa content. *Food Quality and Preference, 22*, 567 – 572.

Nielsen, A., & Rendall, D. (2011). The sound of round: Evaluating the sound-symbolism role of consonants in the classic Takete-Maluma phenomenon. *The Canadian Journal of Experimental Psychology 65* (2), 115-124.

Noppeney, U. (2007). The effects of visual deprivation on functional and structural organization of the human brain. *Neuroscience Biobehavioral Review, 31*,1169 - 1180.

Nunez, D. (2007). Effects of non-diegetic information on presence: A content manipulation experiment. *Proceedings of the 10th Annual International Workshop on Presence, 19-25. October, Barcelona*.

Nunez, D., & Blake. E. H. (2006). Content knowledge and thematic inertia predict virtual presence. *Proceedings of the ninth international workshop on presence (PRESENCE 2006) Cleveland, OH*.

Oberman, L. M., & Ramachandran, V. S. (2008). Preliminary evidence for deficits in multisensory integration in autism spectrum disorders: the mirror neuron hypothesis. *Social Neuroscience*, *3*(3-4), 348-355.

Occelli, V., Spence, C., & Zampini, M. (2008). Audiotactile temporal order judgments in sighted and blind individuals. *Neuropsychologia, 46* (11), 2845-2850.

Ofcom. (2008). *Code on Television Access Services. London: Ofcom*. Retrieved http://www.ofcom.org.uk/static/archive/itc/itc_publications/codes_guidance/audio_description/audio_2.asp.html

Ofcom (2009). *Research into the awareness and usage of Audio Description.* http://www.ofcom.org.uk/research/tv/reports/research_audio_description

Ofcom (2012). *Television access services final report*. Retrieved from http://stakeholders.ofcom.org.uk/market-data-research/market-data/tv-sector-data/tv-access-services-reports/2012-report

Ogden, J. A., & Barker, K. (2001). Imagery used in autobiographical recall in early and late blind adults. *Journal of Mental Imagery 25*, 135 – 152.

O'Keefe, J., & Nadel, L. (1978). *The hippocampus as a cognitive map.* Oxford, England: Clarendon.

Oliver, M. B. (1993). Exploring the paradox of the enjoyment of sad films. *Human Communication Research, 19,* 315–342.

Paivio, A. (1977). *Imagery and verbal processes*. New York: Holt, Rinehart & Winston.

Park, G., Van Bavel, J. J., Vasey, M. W., & Thayer, J. F. (2012). Cardiac vasal tone predicts attentional engagement to and disengagement from fearful faces. *Emotion*, *12*(6), 1292-302.

Pascual-Leone, A., & Hamilton, R. (2001). The metamodal organization of the brain. *Progress in Brain Research*, *134*, 427-45.

Passini, R., & Proulx, G. (1988). Wayfinding without vision: An experiment with congenitally totally blind people. *Environment & Behavior*, 20(2), 227-252.

Peli, E., & Fine, E. (1996). Evaluating visual information provided by audio description. *Journal of Visual Impairment and Blindness, 90* (5), 378 - 385.

Perez Pereira, M., & Conti-Ramsden, G. (2004). Do blind children show autistic features? In L. Pring (Ed.), *Autism and blindness: Research and reflections* (pp. 99–127). London. Wiley.

Pesky People (2011). *National Theatre: Accessibility experts*. Pesky People - Disability meets digital. Retrieved 20 Oct. 2011 from http://www.peskypeople.co.uk/2011/09/national-theatre-accessibility-experts/

Petrie, H., Hamilton, F., King, N., & Pavan, P. (2006). Remote usability evaluations with disabled people. In *Proceedings of the SIGCHI conference on Human Factors in computing systems* (pp. 1133-1141). ACM.

Pettitt, B., Sharpe, K. & Cooper, S. (1996). AUDETEL: Enhancing telesight for visually impaired people. *British Journal of Visual Impairment 14* (2), 48 - 52.

Pietrini, P., Ptito, M., & Kupers, R. (2009). Blindness and consciousness: new lights from the dark. *The Neurology of Consciousness*, 360-374.

Piety, P. (2004). The language system of audio description: an investigation as a discursive process. *Journal of Visual Impairment and Blindness, 98(8)*, 453-469.

Pinchbeck, D. M., & Stevens, B. (2005). Presence, narrative and schemata. In *Proceedings of Presence 2005: The 8th International Workshop on Presence* (pp. 221-26). Retrieved from http://www.temple.edu/ispr/prev_conferences/proceedings/2005/PinchbeckandStevens.pdf

Plantinga, C. R., & Smith, G. M. (1999). *Passionate views: Film, cognition, and emotion*. Johns Hopkins University Press.

Postma, A., Zuidhoek, S., Noordzij, M. L., & Kappers, A. M. (2007). Differences between early blind, late blind and blindfolded sighted people in haptic spatial configuration learning and resulting memory traces. *Perception*, *36*(8), 1253-1265.

Prince, S. E., Daselaar, S. M., & Cabeza, R. (2005). Neural correlates of relational memory: successful encoding and retrieval of semantic and perceptual associations. *Journal of Neuroscience, 25*(5), 1203-1210.

Pring, L. (2002). *Loss and gain: research on blindness, autism and talent*. Inaugural lecture. Goldsmiths College, London.

Pring, L. (2004). *Autism and blindness: Research and reflections*. London: Wiley.

Pulvermüller, F. (2005). Brain mechanisms linking language and action. *Nature Reviews Neuroscience*, *6*(7), 576-582.

Putzar, L., Goerendt, I., Lange, K., Rösler, F., & Röder, B. (2007). Early visual deprivation impairs multisensory interactions in humans. *Nature Neuroscience*, *10*(10), 1243-1245.

Putzar, L., Gondan, M., & Röder, B. (2012). Basic multisensory functions can be acquired after congenital visual pattern deprivation in humans. *Developmental Neuropsychology*, *37*(8), 697-711.

Radio Advertising Bureau (online). *Radio listening habits*. Retrieved from http://www.rab.co.uk/why-use-radio/listener-insight/radio-listening-habits

Rai, S., Greening, J., & Petre, L. (2010). *A comparative study of audio description guidelines prevalent in different countries.* Media and Culture Department, Royal National Institute of Blind People. London.

Ramachandran, V. S., & Hubbard, E.M. (2001). Synaesthesia – a window into perception, thought and language. *Journal of Consciousness Studies, 8* (12), 3-34.

Ramachandran, V. S., & Oberman, L.M. (2006). Broken Mirrors: A Theory of Autism. *Scientific American 295 (5),*62-69.

Raz, N., Striem, E., Pundak, G., Orlov, T., & Zohary, E. (2007). Superior serial memory in the blind: a case of cognitive compensatory adjustment. *Current Biology, 17*(13), 1129-1133.

Rendall, D., Owren, M. J., & Ryan, M. J. (2009). What do animal signals mean? *Animal Behaviour*, *78*(2), 233-240.

Ricciardi, E., & Pietrini, P. (2011). New light from the dark: what blindness can teach us about brain function. *Current Opinion in Neurology*, *24*(4), 357-363.

Richards, I. A. (1972). *Plato's Republic*. Cambridge: Cambridge University Press.

Richardson, J.T.E. (1999). *Imagery*. Hove, UK. Psychology Press Ltd.

Ridgeway, D., & Waters, E. (1987). Induced mood and preschoolers' behavior: Isolating the effects of hedonic tone and degree of arousal. *Journal of Personality and Social Psychology*, *52*(3), 620.

Rieser, J. J., Guth, D. A., & Hill, E. W. (1986). Sensitivity to perspective structure while walking without vision. *Perception*, *15*(2), 173-188.

Rieser, J.J., Lockman, J. L., & Pick, H.L (1980) The role of visual experience in knowledge of a spatial layout. *Perception and Psychophysics 28*, 185-90.

Riva, G., Mantovani, F., Capideville, C. S., Preziosa, A., Morganti, F., Villani, D., Gaggioli, A., Botella, C., & Alcañiz, M. (2007). Affective interactions using virtual reality: The link between presence and emotions. *CyberPsychology & Behavior, 10*(1), 45-56.

Rock, I. (1985). Perception and knowledge. *Acta Psychologica*, *59*(1), 3-22.

Röder, B., Föcker, J., Hötting, K., & Spence, C. (2008). Spatial coordinate systems for tactile spatial attention depend on developmental vision: evidence from event-related potentials in sighted and congenitally blind adult humans. *European Journal of Neuroscience*, *28*(3), 475-483.

Röder, B., & Rösler, F. (2004). Compensatory plasticity as a consequence of sensory loss. *The Handbook of Multisensory Processes*, 719-747.

Röder, B., Rösler, F., Hennighausen, E., & Näcker, F. (1996). Event-related potentials during auditory and somatosensory discrimination in sighted and blind human subjects. *Cognitive Brain Research*, *4*(2), 77-93.

Röder, B., Rösler, F., & Spence, C. (2004). Early vision impairs tactile perception in the blind. *Current Biology*, *14*(2), 121-124.

Rodero, E. (2012). See it on a radio story: sound effects and shots to evoked imagery and attention on audio fiction. *Communication Research, 39*, 458–479.

Rokem, A., & Ahissar, M. (2009). Interactions of cognitive and auditory abilities in congenitally blind individuals. *Neuropsychologia, 47*(3), 843-848.

Romano, D. M., Brna, P., & Self, J. A. (1998, June). Collaborative decision-making and presence in shared dynamic virtual environments. In *Proceedings of the Workshop on Presence in Shared Virtual Environments.* BT Labs, Martlesham Heath.

Rosenblum, L. D., Paige Wuestefeld, A., & Anderson, K.L. (1996). Auditory reachability: An affordance approach to the perception of sound source distance. *Ecological Psychology, 8 (1*), 1 – 24.

Ross, D. A., & Kelly, G. W. (2009). Filling the gaps for indoor wayfinders. *Journal of Visual Impairment and Blindness (April)*, 229 – 234.

Ross, K. (2003). All ears: Radio, reception and discourses of disability. In V. Nightingale, and K. Ross (Eds.), *Critical Readings: Media and Audiences*, (pp. 131-144). Maidenhead, Berkshire: Open University Press.

Roth, I. (1986). An introduction to object perception. In I. Roth & J. P. Frisby (Eds.). *Perception and representation: A cognitive approach*. Milton Keynes: Open University Press.

Royal National Institute of Blind People (2012). *Sight loss UK 2012 – The latest evidence*. Retrieved from http://www.rnib.org.uk/aboutus/Research/reports/otherresearch/Pages/sight_loss_uk.aspx

Russell, B. (1917). *Mysticism and Logic*. London: George Allen & Unwin Ltd.

Sacau, A., Laarni, J., & Hartmann, T. (2008). Influence of human factors on presence. *Computers in Human Behaviour 24*, 2255-2273.

Sacks, O. (2003). The mind's eye. *New Yorker (28),* 48-59.

Sander, D., Grandjean, D., Pourtois, G., Schwartz, S., Seghier, M. L., Scherer, K. R., & Vuilleumier, P. (2005). Emotion and attention interactions in social cognition: Brain regions involved in processing anger prosody. *Neuroimage*, *28*(4), 848-858.

Santangelo, V., & Spence, C. (2008). Is the exogenous orienting of spatial attention automatic? Evidence from unimodal and multisensory studies. *Consciousness and Cognition 17*, 989 – 1015.

Sauerburger, D., & Bourquin, E. (2010) Teaching the use of a long cane step by step: Suggestions for progressive, methodological induction. *Journal of Visual Impairment and Blindness (April)*, 203 – 214.

Schacter, D. L. (1992). Priming and multiple memory systems: Perceptual mechanisms of implicit memory. *Journal of Cognitive Neuroscience 4*(3), 244-256

Schank, R. C., & Abelson, R. P. (1977). *Scripts, plans, goals, and understanding*. Hillsdale, N.J.: Erlbaum.

Schenkman, B. N. (1986). Identification of ground materials with the aid of tapping sounds and vibrations of long canes for the blind. *Ergonomics*, 29, 985-998.

Schmeidler, E., & Kirchner, C. (2001). Adding audio description. Does it make a difference? *Journal of Visual Impairment and Blindness 95*(4), 197-212.

Schubert, T., Friedmann, F., & Regenbrecht, H. (2001). The experience of presence: Factor analytic insights. *Presence: Teleoperators and Virtual Environments, 10(3),* 266–281.

Schrader, C., & Bastiaens, T. J. (2012). The influence of virtual presence: Effects on experienced cognitive load and learning outcomes in educational computer games. *Computers in Human Behavior, 28*(2), 648-658.

Sepchat, A., Descarpentries, S., Monmarche, N., & Slimane, M. (2008). MP3 players and audio games: An alternative to portable video games console for visually impaired players. *Computers helping people with special needs. Lecture notes in computer science* (5105), 553 – 560.

Sgoifo, A., Braglia, F., Costoli, T., Musso, E., Meerlo, P., Ceresini, G., & Troisi, A. (2003). Cardiac autonomic reactivity and salivary cortisol in men and women exposed to social stressors: Relationship with individual ethological profile. *Neuroscience and Biobehavioural Reviews, 27,* 179-188.

Shapiro, M., & McDonald, D. (1993). I'm not a doctor, but I play one in virtual reality. *Journal of Communication, 42* (4), 94-114.

Shedlosky-Shoemaker, R., Costabile, K. A., DeLuca, H. K., & Arkin, R. M. (2011). The social experience of entertainment media. *Journal of Media Psychology: Theories, Methods, and Applications*, *23*(3), 111-121.

Sheehan, P.W. (1967). A shortened form of Bett's Questionnaire Upon Mental Imagery (QMI). *Journal of Clinical Psychology*, 23, 386-389.

Sheridan, T. B. (1992a). Musings on telepresence and virtual presence. *Presence: Teleoperators and Virtual Environments,1,* 120–125.

Shu, N., Liu, Y., Li, J., Li, Y., Yu, C., & Jiang, T. (2009). Altered anatomical network in early blindness revealed by diffusion tensor tractography. *PLoS One*, *4*(9), e7228.

Skalski, P., & Whitbred, R. (2010). Image versus sound: A comparison of formal feature effects on presence and video game enjoyment. *PsychNology Journal, 8 (1)*, 67 – 84.

Slater, M. (1999). Measuring presence: A response to the Witmer and Singer presence questionnaire. *Presence:Teleoperators and Virtual Environments, 8*, 560-565

Slater, M., (2002). Presence and the sixth sense. *Presence: Teleoperators and Virtual Environments, 11*(4), 435-439.

Slater, M., & Usoh, M. (1993). Representation systems, perceptual positions, and presence in immersive virtual environments. *Presence 2,* 221-233.

Slater, M., Usoh, M., & Steed, A. (1994). Depth of presence in virtual environments. *Presence: Teleoperators and Virtual Environments, 3*, 130-144.

Smith, E. E., & Kosslyn, S. M., (2009). *Cognitive psychology: Mind and brain*. Upper Saddle River, NJ: Pearson Prentice Hall.

Smith, G. M. (1999). Local emotions, global moods and film structure. In C. Plantinga & G. M. Smith (Eds.). *Passionate views: film, cognition and emotion* (pp. 103 – 126). Baltimore: John Hopkins University Press.

Smith, T. J. (2006). *An attentional theory of continuity editing*. (Doctoral dissertation, University of Edinburgh). Retrieved from http://hdl.handle.net/1842/1076

Smith, T. J. (2011, February). *A trick of the eye: Investigating the dynamics of real-world attention and awareness*. Department of Psychology seminar given at Goldsmiths University, London.

Smith, T. J., Levin, D. T., & Cutting, J. (2012). A Window on reality: Perceiving edited moving images. *Current Directions in Psychological Science. 21*, 101-106 doi:10.1177/0963721412436809.

Spence, C. (2012). Managing sensory expectations concerning products and brands: Capitalizing on the potential of sound and shape symbolism. *Journal of Consumer Psychology, 22*, 37-54.

Spence, C., & Gallace, A. (2011a). Tasting shapes and words. *Food Quality & Preference, 22*, 290-295.

Spence, C., & Gallace, A. (2011b). Multisensory design: Reaching out to touch the consumer. *Psychology & Marketing, 28*, 267 - 308.

Stanney, K. M., Mourant, R. R., & Kennedy, R. S. (1998). Human factors issues in virtual environments: A review of the literature. *Presence: Teleoperators and Virtual Environments, 7*(4), 327–351.

Stern, R. M., Ray, W. J., & Quigley, K. S. (2001). *Psychophysiological Recording*. New York: Oxford University Press.

Steuer, J. (1992). Defining virtual reality: Dimensions determining telepresence. *Journal of Communication, 4* (*2*4), 73-93.

Steuer, J. (1995). Defining virtual reality: Dimensions determining telepresence, in F. Biocca and M.R. Levy (Eds.), *Communication in the Age of Virtual Reality*. Lawrence Erlbaum.

Stevens, A. A., & Weaver, K. E. (2009). Functional characteristics of auditory cortex in the blind. *Behavioural Brain Research, 196*, 134–138.

Struiksma, M. E., Noordzij, M. L., & Postma, A. (2009). What is the link between language and spatial images? Behavioral and neural findings in blind and sighted individuals. *Acta Psychologica, 132*(2), 145-156.

Struiksma, M. E., Noordzij, M. L., & Postma, A. (2011). Reference frame preferences in haptics differ for the blind and sighted in the horizontal but not in the vertical plane. *Perception, 40*, 725 – 738.

Sutcliffe, T. (2000). *Watching*. London: Faber & Faber.

Sweller, J. (1988). Cognitive load during problem solving: Effects on learning. *Cognitive science, 12*(2), 257-285.

Sweller, J. (2010). Element interactivity and intrinsic, extraneous, and germane cognitive load. *Educational Psychology Review, 22*(2), 123-138.

Szalwinska, M. (2009). '*You feel swept up by it': the drama of audio description*. The Guardian 24 Feb, 2009. Retrieved from http://www.guardian.co.uk/stage/2009/feb/24/drama-audio-description

Szarkowska, A., & Jankowska, A. (2011). Text-to-speech audio description: Towards wider availability of AD. *Journal of Specialised Translation 15* (1), 81 – 98.

Tabakowska, E. (2003). Iconicity and literary translation. In W.G. Muller & O. Fischer (Eds.), *From sign to signing: Iconicity in language and literature 3* (pp. 361 – 378). Amsterdam, Philadelphia: Benjamins.

Tamborini, R., & Schiff, J. (1987). Predictors of horror film attendance and appeal: An analysis of the audience for frightening films. *Communication Research, 14*, 415–436.

Thagard, P, (2012). Cognitive Science. In Edward N. Zalta (Ed). *The Stanford Encyclopedia of Philosophy*. Retrieved from <http://plato.stanford.edu/archives/fall2012/entries/cognitive-science/>.

Théoret, H., Merabet, L., & Pascual-Leone, A. (2004) Behavioral and neuroplastic changes in the blind: Evidence for functionally relevant cross-modal interactions. *Journal of Physiology, 1,* 221-233.

Thomas, N. J. (1999). Are theories of imagery theories of imagination?: An active perception approach to conscious mental content. *Cognitive Science*, *23*(2), 207-245.

Thompson-Schill, S. L. (2003). Neuroimaging studies of semantic memory: inferring "how" from "where". *Neuropsychologia*, *41*(3), 280-292.

Torralba, A., Aude, O., Castelhano, M. S., & Henderson, J. M. (2006). Contextual guidance of attention in natural scenes: the role of global features on object search. *Psychological Review, 113*, 766–786.

Ulrich, I., & Nourbakhsh, I. (2000, July). Appearance-based obstacle detection with monocular color vision. In *Proceedings of the National Conference on Artificial Intelligence* (pp. 866-871). Menlo Park, CA; Cambridge, MA; London; AAAI Press; MIT Press; 1999.

Ungar, S. (2000). Cognitive mapping without visual experience. In R. Kitchin & S. Freundschuh (Eds.), *Cogntive Mapping: Past, Present and Future* (pp. 221-

248). London: Routledge.

Ungar, S., Blades, M., & Spencer, C. (1996). The construction of cognitive maps by children with visual impairments. In J. Portugali (Ed.), *The Construction of Cognitive Maps* (pp. 247 – 273). Dordrecht: Kluwer Academic Publishers.

Ungar, S., Simpson, A., & Blades, M. (2004). Strategies for organising information while learning a map by blind and sighted people. In M. Heller & S. Ballasteros (Eds.), *Touch, Blindness and Neuroscience*. Madrid: Universidad Nacional de Educacion a Distancia.

Västfjäll, D. (2003). The subjective sense of presence, emotion recognition and experienced emotions in auditory virtual environments. *Cyberpsychology & Behaviour, 6*(2), 181-188.

Velmans, M. (1999). When perception becomes conscious. *British Journal of Psychology, 90*, 543 – 566.

Velmans, M. (2008). Reflexive monism. *Journal of Consciousness Studies* (pre-publication version now published as Reflexive monism: Psychophysical relations among mind, matter, and consciousness). J*ournal of Consciousness Studies 2012, 19* (9-10), 143 – 165.

Vilaró, A., & Orero, P. (2013). Leitmotif in audio description: Anchoring information to optimise retrieval. *International Journal of Humanities and Social Science, 3*(5), 56 - 64.

Vogeley, K., & Fink, G. R. (2003). Neural correlates of the first-person-perspective. *Trends in Cognitive Sciences*, *7*(1), 38 - 42.

Voss, P., Lassonde, M., Gougoux, F., Fortin, M., Guillemot, J. P., & Lepore, F. (2004). Early-and late-onset blind individuals show supra-normal auditory abilities in far-space. *Current Biology*, *14*(19), 1734 -1738.

Wall Emerson, R. S., & Corn, A. L. (2006). Orientation and mobility content for children and youths: A Delphi approach pilot study. *Journal of Visual Impairment and Blindness (June)*, 331 – 341.

Waller, D., & Greenauer, N. (2007). The role of body-based sensory information in the acquisition of enduring spatial representations. *Psychological Research*, 71, 322 - 332. DOI 10.1007/s00426-006-0087-x

Wan, C. Y., Wood, A.G., Reutens, D. C., & Wilson, S. J. (2010). Early but not late blindness leads to enhanced auditory perception. *Neuropsychologia*, Jan (48), 344-348.

Ward, R. D., Marsden, P. H., Cahill, B., & Johnson, C. (2002). Physiological responses to well-designed and poorly designed interfaces. In *Proceedings of CHI 2002 Workshop on Physiological Computing*. Minneapolis, MN, USA.

Waterworth, J., & Waterworth, E. (2001). Focus, locus and sensus: The 3 dimensions of virtual experience. *CyberPsychology and Behaviour 4* (2), 203-214.

Waterworth, J., & Waterworth, E. (2003a). The meaning of presence. *Presence-Connect 3.* Retrieved from http://www8.informatik.umu.se/~jwworth/PRESENCE-meaning.htm

Waterworth, J., & Waterworth, E. (2003b). Being and time: Judged presence and duration as a function of media form. *Presence: Teleoperators and Virtual Environments 12* (5), 495 – 511.

Whitehead, J. (2005). What is audio description? *International Congress Series 1282*, 960 – 963.

Weis, E. (1982). *The silent scream: Alfred Hitchcock's sound track*. Fairleigh Dickinson University Press.

Wierzbicka, A. (1991). *Cross-cultural pragmatics*. Berlin, New York: Mouton de Gruyter.

Wiethoff, S., Wildgruber, D., Kreifelts, B., Becker, H., Herbert, C., Grodd, W., & Ethofer, T. (2008). Cerebral processing of emotional prosody - influence of acoustic parameters and arousal. *Neuroimage*, *39*(2), 885-893.

Willander, J., & Baraldi, S. (2010). Development of a new clarity of auditory imagery scale. *Behavior Research Methods, 42* (3), 785-790.

Williams, J. M. G., Healy, H. G., & Ellis, N. C. (1999). The effect of imageability and predictability of cues in autobiographical memory. *The Quarterly Journal of Experimental Psychology*, 52A, 555-579.

Wilson, G. M., & Sasse, M. A. (2000). Do users always know what's good for them? Utilising physiological responses to assess media quality. In *People and Computers XIV—Usability or Else!* (pp. 327-339). Springer London.

Winawer, J., Witthoft, N., Frank, M. C., Wu, L., Wade, A. R., & Boroditsky, L. (2007). Russian blues reveal the effects of language on colour discrimination. *Proceedings of the National Academy of Sciences of the United States of America, 104* (19), 7780 – 7785.

Wirth, W., Hartmann, T., Bocking, S., Vorderer, P., Limmt, C., Schramm, H., Saari, T., Laarni, J., & Ravaja, N. (2007). A process model of the formation of spatial presence experiences. *Media Psychology, 9*, 493 – 525.

Wissmath, B., Weibel, D., & Groner, R. (2009). Dubbing or subtitling?: Effects on spatial presence, transportation, flow, and enjoyment. *Journal of Media Psychology: Theories, Methods, and Applications*, *21*(3), 114 – 125.

Woods, A., & Newell, T. (2004). Visual, haptic and cross-modal recognition of objects and scenes. *Journal of Physiology, Paris. 98(1-3),* 147 – 159.

Worsfold, J. & Chandler, E. (2010) *Wayfinding Project Final report*. London: RNIB.

Wyer, R. S., Hung, I. W., & Jiang, Y. (2008). Visual and verbal processing strategies in comprehension and judgment. *Journal of Consumer Psychology, 18*, 247 - 257.

Yoo, S. S., Lee, C.U., & Choi, B. G. (2001). Human brain mapping of auditory imagery: event-related functional MRI study. *NeuroReport, 12,* 3045 – 3049.

Zabrocka-Suwka, M. (2013, September). *Audio description as a tool for creating the linguistic image of the world by blind children*. Paper presented at the 5th Media for All Conference, Dubrovnik.

Zacks, J. M. (2010). How we organize our experience into events. *Psychological Science Agenda, 24*. Retrieved from http://www.apa.org/science/about/psa/2010/04/sci-brief.aspx

Zahorik, P., & Jenison, R. L. (1998). Presence as being-in-the-world. *Presence 7 (1)*, 78-89.

Zangenehpour, S., & Zatorre, R.J. (2009). Crossmodal recruitment of primary visual cortex following brief exposure to bimodal audiovisual stimuli. *Neuropsychologia*, *48*, 591- 600.

Zatorre, R. J., Halpern, A. R., & Bouffard, M. (2010). Mental reversal of imagined melodies: a role for the posterior parietal cortex. *Journal of Cognitive Neuroscience*, *22*(4), 775 - 789.

Zelter, D. (1992). Autonomy, interaction, and presence. *Presence: Teleoperators and Virtual Environments 1* (1), 1275 – 132.

Zhou, W., Jiang, Y., He, S., & Chen, D. (2010). Olfaction modulates visual perception in binocular rivalry. *Current Biology, 20*(15), 1356 -1358.

Zillmann, D. (2003). Theory of affective dynamics: Emotions and moods. In J. Bryant, D. Roskos-Ewoldsen, & J. Cantor (Eds.), *Communication and emotion: Essays in honor of Dolf Zillmann* (pp. 533–567). Mahwah, NJ: Lawrence Erlbaum Associates.

Zimler, J., & Keenan, J. M. (1983). Imagery in the congenitally blind: How visual are visual images? *Journal of Experimental Psychology: Learning, Memory and Cognition, 9*, 269 - 282.

Zuidhoek, S. (2005). *Representations of space based on haptic input*. Doctoral thesis from the University of Utrecht.

Filmography

Bender, L. (Producer), & Tarantino, Q., & Roth, E. (Directors) (2009). *Inglourious basterds* [Motion Picture]. USA: Universal Pictures.

Broadbent, G., Hauptman, A., & Jones, D. (Producers) & Boyle, D. (Director) (2004). *Millions* [Motion Picture]. UK: Pathe.

Cooper, R. (Producer), & Minghella, A. (Director) (1990). *Truly, madly, deeply* [Motion Picture]. UK, London: BBC Films.

Disney, W. (Producer), & Hand, D. (Director) (1942). *Bambi* [Motion Picture]. USA, Burbank, CA: Walt Disney Home Video.

Kubrick, S. (Producer and Director) (1980). *The Shining* [Motion Picture]. USA, Burbank, CA: Warner Home Video.

Myers, M., Moore, D., Todd, J., & Todd, S. (Producers) & Roach, J. (Director). (1997). *Austin Powers: International man of mystery.* [Motion picture]. United States: New Line Cinema.

Kenworth, D. (Producer) & Newell, M. (Director) (1994). *Four weddings and a funeral* [Motion Picture]. UK, London.

Parkes, W. F., & MacDonald, L. (Producers), & Verbinski, G. (Director) (2002). *The Ring* [Motion Picture]. USA, California: Dreamworks Pictures.

Saltzman, H., & Broccoli, A. (Producers), & Young, T. (Director) (1963). *To Russia with love*. [motion picture]. United States: United Artists.

Saxon, E., Utt, K., Bozman, R. (Producers), & Demmie, J. (Director) (1991). *Silence of the lambs* [Motion Picture]. USA, New York, NY: Orion Pictures.