PERCEPTUAL LATERALISATION OF AUDIO-VISUAL STIMULI

By

Nigel James Holt

Dissertation submitted to the University of York for the Degree of

Doctor of Philosophy

September 1997

Department of Psychology

University of York

CONTENTS

CHAPTER 2

CHAPTER 3

CHAPTER 4

CHAPTER 5

ACKNOWLEDGEMENTS

I would like to thank Tony Watkins of the University of Reading for making auditory psychophysics appealing, and for encouraging me to undertake this study.

Many thanks to my supervisor Peter Bailey, without whose help and encouragement I would not have completed this thesis. His comments on draft copies have been invaluable and his willingness to read and discuss work has been far beyond the call of duty.

Thanks goes to the technical staff for recognising the urgency with which I often required leads and hardware repairs. Thanks too to Rob Stone, whose assistance with the code for the early experiments saved much time and agony. I also thank my long suffering subjects who sat through many hours of experiments. Thanks to Nick Hill for the use of his computer.

I must thank my parents for all their support. Their help, both personal and material has been invaluable as always.

Thanks to Jack and Lewis, for helping me to realise that there's more to life than psychophysics.

Finally, thanks to Kate. Her support and encouragement have been limitless. The last year has been a seemingly interminable haul, and without her I shudder to think what would have become of me.

ABSTRACT

This thesis is concerned with the perceptual integration of auditory and visual stimuli. Lateralisations of auditory, visual and audio-visual stimuli involving simple non-speech sounds were investigated. A correspondence between visually and auditorily presented lateral positions was demonstrated, allowing the presentation of audio-visual stimuli with laterally corresponding modal components. Mean lateralisation judgements of auditory, visual and audio-visual stimuli with spatio-temporally corresponding components did not differ significantly. Mean standard deviations in lateralisation judgements of audio-visual stimuli with spatio-temporally corresponding components were significantly smaller than those of judgements of auditory or visual stimuli suggesting that subjects' lateralisation judgements were not based solely on stimulus properties of one or other modal component. Measurements of thresholds for the detection of audio-visual spatial mismatch provided normative data for the assessment of lateralisation judgements of audio-visual stimuli with spatially non-corresponding components. A dominance of the visual modality was shown (c.f. Radeau and Bertelson 1977), but mean standard deviations in lateralisation judgements increased as a function of audio-visual spatial mismatch. Audio-visual temporal mismatch difference limen measurements suggested that auditory processing time was approximately 50ms faster than visual processing time (c.f. Poppel 1988). Lateralisation judgements of audio-visual stimuli with asynchronous auditory and visual components showed an increase in standard deviation as the asynchrony increased. Mean lateralisation judgements of audio-visual stimuli with spatio-temporally non-corresponding components showed an influence of the position of the auditory component. The relative influence of the auditory component was attributed to increased stimulus unpredictability as a result of the simultaneous variation of audio-visual spatial and temporal correspondence. The experiments are discussed in terms of the influences of structural and cognitive variables on the perception that the individual modal components refer to the same perceptual event - the assumption of unity. Models of cross-modal perceptual integration are discussed and areas for further study suggested.

Chapter 1

1.1 INTRODUCTION

"The act of interpreting a stimulus, registered in the brain by one or more sense mechanisms" [1]

This definition of perception is attractively simple. A stimulus causes a change of some kind in one or more sense organ, initiating electrical and chemical differences within the brain, carrying the information to the appropriate neural centers. The organism responds on the basis of calculations made. The response is usually appropriate and accurate.

How an infant acquires the mechanisms that provide these perceptions is interesting in itself, and some of the developmental evidence will be discussed in this review. The reference to changes in one or more sense organ caused by a single stimulus is of direct interest to this study. A stimulus that is capable of stimulating more than one sense organ is usually a multi-modal stimulus. Combinations of smell, heat, light and sound can be registered as a single perceptual event rather than a combination of stimuli in different modalities.

[1] Psychology (p.268) Sperling A., Martin K., Heinemann 1986.

For this multi-modal stimulus to be interpretable as a single stimulus a decision needs to be made regarding whether the information, carried separately by the two or more modalities, belongs together. This phenomenon is described by Radeau and Bertelson (1977) as the Assumption Of Unity (AOU). Evidence of the neural basis for this multi-sensory integration is also reviewed here.

This thesis is concerned with one specific type of multi-modal stimulus - a non-speech audio-visual stimulus - and how perception of the stimulus is altered when the component modalities carry conflicting information. This review also includes literature concerned with perception of non-audio-visual combinations and audio-visual speech, where the auditory component, heard words, is linked with the visual component, moving lips.

1.2 DEVELOPMENT OF INTERSENSORY INTEGRATION

"The multi-modal perception of an object is the ability to perceive different pieces of information extracted by the sensory modalities in a unified way" [1]

The evolutionary advantages of this intersensory integration are obvious. Stein and Meredith (1993) show that if the location of an animal is given by

[1] p.285 Streri and Molina (1994) Constraints on intermodal transfer between touch and vision in infancy. In Lewkowicz and Lickliter eds. (1994)

both auditory and visual cues, a predator will be more accurate in locating it than when only uni-modal information is available.

It can be argued that not having the ability to integrate information from its several senses threatens a child's survival in a multi-modal environment. Social and communicative skills would both be impaired if sensory integration were not possible. These points will be discussed later.

Whether sensory integration skills are innate is still a point of some disagreement. Explanations of their origin fall into two camps, adding to the much-argued nature/nurture debate. Piaget (1954) proposes that the child's activities give rise to perceptual experiences which strengthen the link between two or more senses. For instance if a child is given a ball, inside which is a bell, the characteristic sound and visual appearance of the toy will be present each time it is manipulated. Repetition of this co-occurrence of multi-sensory stimulation serves to reinforce audio-visual cross-modal links. On the other hand, Gibson (1979) suggests that the amodal elements of the object - those perceptual properties common to sound and vision, e.g. location - in themselves allow cross-modal matching. The very fact that the location characteristics are amodal is in itself enough to link the two separate sensory streams in a representation of a unitary stimulus in which the amodal elements are functionally equivalent.

The cross-modal abilities of infants that have been investigated in detail can be divided roughly in two: touch and vision, and sound and vision.

1.2.1 TOUCH AND VISION

Streri and Molina (1994) report evidence showing that intermodal transfer between touch and vision is possible after the age of 6 months. An intermodal transfer matching technique was used which assessed the child's recognition of the identity of an object previously experienced in one modality from stimuli presented in another. The ability of six month olds to do this, say Streri and Molina, provides evidence to support the existence of a mechanism independent of the modality receiving the informative stimulus. Other studies (Meltzoff and Borton 1979) have shown transfer in infants as young as 1 month, although anomalies were found. Intermodal transfer, for instance, can be shown between information obtained visually and tested for haptically, but not between information obtained haptically and tested for visually. Further investigation from Streri & Pêcheux (1986) suggested that this may be due to the infant's inability to extract perceptually important information from touch until around 5 months. However, they concluded that the inability to transfer information about an object from touch to vision was more likely to be a function of the infants' poor motor ability at that age. 5 month old infants receive more stimulation visually and by placing objects in the mouth than they do by simply manipulating them. Streri and Molina (1994) suggest that the lack of touch to vision transfer is because the infant perceives a held object as one to be looked at and sucked rather than to be recognised. On balance,

there is agreement with Piaget's theory of vision prehension as occurring at around four and a half months. Although it is accepted (Bloch 1990, Bower 1979) that some links between vision and touch are possible at even a few days old, this is not evidence for any prolonged relationship between the two senses (Bloch 1994). Both Bloch (1994) and Streri and Molina (1994) agree that intermodal transfer between touch and vision and vice-versa is dependent on the infant's ability to extract perceptually salient, potentially amodal information using both senses. Until five and a half to six months the haptic sense has not developed enough for free transfer of information between the senses in both directions (Bloch 1994; Streri and Molina 1994; Streri and Pêcheux 1986).

1.2.2 SOUND AND VISION

The interaction between sound and vision provides a rich source for developmental research. The amodal cues to an object's identity provided by both senses aid the infant in both its communication and physical interactions with the world. It is useful to divide this evidence into non-speech audio-visual, and speech-related audio-visual interactions.

1.2.2.a Non-speech audio-visual interactions.

Morrongiello (1994) reports that visual attention in infancy is increased by providing a sound in the same location as the visual target. She goes on to note that the properties of the sound mediate the kind of response. Loud sounds encourage eye movement away from the target, soft sounds towards it.

This co-location effect suggests an audio-visual interaction based on the nature of the visual and auditory information available for locating the stimulus, in that the information in both modalities identifies the stimulus' location. Whereas the relevance of the differential effects of the physical properties of the sound is unclear, they indicate that the nature of the sound is also a factor in the strength of intermodal interactions of this kind.

Bower (1979) cites evidence of a new-born only seconds old (Wertheimer 1961) who reliably turned her eyes in the direction of a sound. This action, says Bower, demands two things:

(i)The ability to locate sound.

(ii) The expectation that there will be something to see at its source.

It seems more likely that the infant was performing a neo-natal orienting reflex (O.R.) facilitated by the 'functionally equivalent' (Gibson 1979) location cue provided by the auditory stimulus, rather than exhibiting an innate understanding of the concept of sound-producing visual objects. This view is supported by evidence cited by Morrongiello (1994) which indicates that the O.R. is not dependent on the presence of a visual object. Infants show orientation to sound sources in the dark, and even with their eyes closed. Piaget (1952) states that at birth the senses are separate, and the child must manipulate its environment to determine how the apparently distinct streams of sensory information relate to one another. The responses of Wertheimer's subject would suggest that this is not the case, and some kind of intermodal

relationship is present at birth. Bower (1982) points out that the O.R. is likely to be innate to "..guarantee experiences that may be crucial to promote perceptual learning and development." Reflexively orienting to a sound source provides the infant with numerous instances of audio-visual temporal and spatial consistency, as well as experience of using one part of the bi-modal stimulus as a cue to the identity of the audio-visual object.

An interesting difference between adult and infant audio-visual integration has been shown by Spelke, Born & Chu (1983). In their experiment two moving objects were linked with one percussive sound. The sound occurred when one of the objects changed direction abruptly, moved through a particular spatial region, or made contact with a rigid surface. Infants showed reactions to the audio-visual stimulus when it changed direction regardless of any impacts, whereas adults responded to the audio-visual relationship only when the object made impact with a surface. This suggests that the infant's perception of an audio-visual relationship depends partly on the detection of a change in movement. The perceptual learning process of non-speech audio-visual correspondence continues well after the child reaches four months of age. As the infant learns and matures, the emphasis shifts to the potentially important link between impacts and sounds.

1.2.2.b Speech-related audio-visual interactions.

Infants show a preference for the human face when presented with an array of pictures (Langsdorf, Izard, Rayais & Hembree 1983). Perhaps the most salient

audio-visual relationship is that of a face and a voice. The biological significance of the pairing is indicated by Bower (1979), reporting an experiment by Aronson and Rosenbloom (1971). When a mother's voice was displaced from her moving mouth infants younger than three weeks showed considerable signs of distress. When the visual and auditory stimuli corresponded spatially, the infants attended to their mothers contentedly. Aronson and Rosenbloom (1971) showed that this distress reaction was no longer shown by infants older than three weeks. Depending on the nature and saliency of the audio-visual relationship, the spatial separation between the modal components can be considerable before adults even notice it (Jackson 1953, Radeau and Bertelson 1977). It seems that neonates put great importance on the spatial correspondence between sound and vision, more than older humans, supporting Bower's (1982) theory of a degree of neonatal sensory correspondence.

Dodd (1979) presented infants with audio-visual stimuli made up of moving lips and corresponding speech. The soundtrack could be in synchrony with the lips, or out of synchrony by 400ms. Her results showed that the ten to sixteen week old infants preferred the synchronous audio-visual stimulus. Piaget (1952) has reported infant imitation of facial gestures, indicating that the relationship between speech and the corresponding visual stimulus is amodal, and has suggested that repetitive imitation helps the infant to learn communication skills. Meltzoff and Kuhl (1994) cite evidence showing that neonates as young as 42 minutes old showed imitation of this kind. The

authors use this finding in support of Meltzoff and Moore's active intermodal mapping (A.I.M.) hypothesis which proposes that infants map their motor output (facial gestures) onto a visual stimulus, in this case the face of the adult. Kuhl and Meltzoff (1982) presented infants with two faces, each paired with a spatially and temporally coincident soundtrack. Only in one of the pairings did the auditory component of the audio-visual stimulus match the visual component so that the face exhibited the correct gestures for the words presented. Infants as young as eighteen to twenty weeks showed a preference for the matching pair, suggesting an understanding of the correct audio-visual relationship. Walton and Bower (1993) showed similar results. They used words from foreign dialects and showed infant preferences for audio-visual stimuli in which facial gestures matched the characteristics of the corresponding word.

Finally, Dodd (1972,1987) showed a relationship between type of stimulation and the amount of vocalisation produced by the infant. Babbling sounds, simulated by an adult, were presented on their own (auditory), or linked with a corresponding facial stimulus (audio-visual). In a 'social' condition the babbling stimulus was presented in a normal mother-child play situation, with the mother instructed to remain silent throughout the session. Results showed that the infant's babbling patterns were altered (increased) only by the audio-visual stimulus. This suggests the importance of this particular audio-visual interaction in the development infants' communication skills.

1.2.3 SUMMARY

Infants have the ability to perceive some intermodal interactions from birth. Wertheimer (1961) showed an orienting reflex in a subject only seconds old. Experiments with interactions between information obtained visually or haptically have indicated the importance of amodal, functionally-equivalent characteristics of stimuli. Dodd (1987, 1979) and Morrongiello (1994) have shown the biologically significant interaction between lips and voices in very young infants. A preference for moving lips with corresponding auditory stimuli over lips paired with non-corresponding stimuli (Meltzoff and Kuhl 1992, Walton and Bower 1993) suggests a link between visually perceived facial gestures and heard speech. Dodd (1979) showed that audio-visual stimuli improved the infants' spontaneous babbling. These results highlight the importance of audio-visual integration in the development of communication and social skills.

1.3 EVIDENCE OF MULTIMODAL NEURAL CENTERS

"There are many areas in the mammalian brain where inputs from two or more sensory systems converge on a single neuron, thereby rendering them multi-sensory."[1]

Several neural areas have neurons responsive to the spatial location of a stimulus. Knudsen and Konishi (1978) showed fields of neurons in the barn owl's MLD (nucleus mesencephalicus lateralis dorsalis) which were responsive to auditory stimuli presented in a specific spatial location. They went on to show that these receptive fields make up a tonotopic map of auditory space. A map of auditory space has been shown in the superior colliculus of the Guinea Pig (King and Palmer 1985) This is consistent with an earlier finding of Knudsen (1983) who showed a similar map in the optic tectum of the owl, the avian homologue of the mammalian superior colliculus (SC). They showed that the map in the optic tectum was provided by projections from the analogous map in the MLD (Knudsen and Konishi 1978).

Several neural areas have neurons responsive to inputs from more that one sensory system (Stein and Meredith, 1993). The superficial layers of the SC receive only visual input, whereas inputs from the auditory, visual and somatosensory systems converge on neurons in the deep laminae area of the

superior colliculus, facilitating the integration of multisensory stimuli (Stein, Meredith, & Wallace, 1994). It is thought that attention and orientation to audio-visual stimuli are influenced by the superior colliculus (Stein 1984, Rauschecker and Kniepert 1993; King and Carlile 1993, Withington-Wray, Binns & Keating 1990). In the cat, the spatial relationship between the stimuli is a primary factor in determining the level of activation of these multimodal neurons (Meredith and Stein 1983, 1986). The cells are activated if the components of the multisensory stimuli are spatially coincident, but depressed or not activated at all if spatially non-coincident audio-visual stimuli are presented. Recordings from single cells showed that combined audio-visual stimuli with spatially corresponding auditory and visual components enhanced the activity of these multi-sensory cells by up to 1207%. Meredith and Stein (1986) showed evidence of a multi-sensory excitatory region and inhibitory region. The firing rate of the cell was increased dramatically if the auditory component of the audio-visual stimulus fell within the excitatory field of the neuron (c.f. Meredith and Stein 1983). When moved outside the excitatory region, the enhancement of cell activity was lost, and if the auditory component fell within the cells' inhibitory region firing rate was depressed. They go on to show that visual, auditory and somatosensory topographic maps are found in the superior colliculus. These maps are overlaid, creating a multisensory space map (Stein et al 1994).

---

[1] P.83 Stein et al (1994). Development and neural basis of multisensory integration. in Lewkowicz & Lickliter eds. 1994

The alignment of this multisensory map is sensitive to manipulation of an animal's sensory inputs early in life. Knudsen's experiments with barn owls (1983, 1989) show that the auditory map is aligned with the visual map. By altering the owl's visual field using prisms in its early development period, a corresponding shift in the auditory map is seen. Similarly, if guinea-pigs are raised in darkness the auditory map is not formed because of the removal of the visual 'template' (Withington-Wray et al. 1989, 1990). Animals raised with one ear plugged also show misalignment of the auditory and visual space maps (King et al 1988). While the plug is in place, the auditory and visual maps appear to be normally aligned. Stein et al (1994) suggest that the brain uses the visual input to weight the relative inputs from the ears to ensure audio-visual alignment. When the plug is removed during the animals' infancy the auditory and visual maps become misaligned, but the auditory map eventually re-aligns with the visual map. If the plug is removed after the animal has reached adulthood, the auditory map remains permanently misaligned. This suggests a period of plasticity, after which the alignment of the maps is fixed (c.f. King, Hutchins, Moore & Blakemore, 1988) . Stein et al (1994) suggest that the auditory map is more plastic than the other maps in the SC as a function of its derivation. The visual connections are direct spatial projections of the topographic map in the retina via the upper layers of the SC to the deep laminae layers of the SC. The auditory map, on the other hand, is calculated in terms of inter-aural cross-reference, and as such must be computed rather than received as a direct projection from the receptors.

Similar evidence of multi-sensory areas have been discovered in primates. Watanabe and Iwai (1991) showed neurons in the ventral intraparietal area in the rhesus monkey which responded to both visual and somatosensory inputs (c.f. Duhamel et al 1989). Visuo-somatosensory and audio-somatosensory cells have also been found in the superior temporal sulcus of the rhesus monkey (Stein and Meredith 1993).

Jay and Sparks (1984) have made recordings from the lower layers of the Monkey SC which indicate that the auditory components of audio-visual signals are changed to take the position of the eyes in the sockets into consideration. Occulocentric spatial information available to the eyes is partly a function of the eyes moving within their orbits. The spatial information available to the auditory system is craniocentric - in relation to the head. Lewald and Ehrenstein (1996) indicate that in order to facilitate the representation of an audio-visual stimulus with spatially corresponding auditory and visual components, the craniocentric co-ordinates of the auditory component should be converted into occulocentric co-ordinates. Jay and Sparks (1984) suggest that the conversion takes place in the lower levels of the SC.

Magnetic Source Imaging (MSI) shows a region of the human cortex which is sensitive to audio-visual stimuli located near the auditory cortex (Regan, He and Regan 1995). Stein, Meredith & Wallace.(1994) have found a similarly placed area in the cat. Unlike the superior colliculus, cells in this region are

sensitive to visual and auditory stimuli. Common spatial source is not necessary for their stimulation, suggesting a specific neural representation of other aspects of audio-visual correspondence, not just common spatial source.

Auerbach and Sperling (1974) assessed the hypothesis that there exists a 'common space' for auditory and visual signals rather than separate auditory and visual representations of the position of an object. In a psychophysical procedure, subjects were presented with an auditory or visual target in one of two locations followed by an ISI. A second signal, either auditory or visual, was presented in one of the two positions. Subjects indicated whether the two stimuli were in the same position or different positions. If both stimuli were visual, performance on the task approached 100%. If the modalities of the two signals differed (mixed-modality trials), performance was less accurate, but there was no difference between the distribution of responses on mixed-modality trials and the distribution of responses on trials in which the two stimuli were auditory, suggesting that the signals referred to a common spatial representation. Auerbach and Sperling concluded that this was evidence of a common auditory and visual representation of spatial location. However, there is the possibility that the common space is a visually based space, with auditory perceptions of space being mapped onto a visually based internal spatial representation. It is true, however that neural evidence has suggested that the superior colliculus contains a multi-modal map of space, providing evidence for an area where visual and auditory representations of space, are in a sense, common.

1.4 GROUPING STRATEGIES

" Our perceptions are unitary. Sights, sounds and the haptic feel of things are coordinated . Thus, we have the central theoretical problem.... How are the separate and qualitatively distinct modalities coordinated and put together?"[1]

Characterising the parsing of elements within the 'separate and qualitatively distinct' modalities provides a framework for understanding the seemingly more complex parsing of multi-modal situations. Understanding how the perceptual environment is parsed is made simpler with reference to the Gestalt principles of perceptual organisation which were developed with reference to the law of Prägnanz.

> "Of several geometrically possible organisations that one will actually occur which possesses the best, simplest and most stable shape."[2]

The principles are diagrammatic, descriptive analogies of essentially unimodal situations, but they are also relevant in a multi-modal context.

Stimuli are often parsed as a function of the dynamics of the scene. In a complicated orchestral piece, individual instruments can often be 'heard out' particularly well if they are producing amplitude or frequency modulations

---

[1] Linda B. Smith page ix. Lewkowicz & Lickliter eds. (1994)

such as vibrato or tremolo (Moore, 1989). Elements are streamed partly on the basis of "common fate", in that they notionally move together. Similarly, the "continuity" principle describes a situation where elements continuing smoothly on from one another in the same direction are more likely to be streamed together than elements whose trajectory is very different.

> "If different parts of the spectrum change in the same way at the same time, they probably belong to the same environmental sound."[1]

The coincidental spectral and temporal changes also increase the "similarity" of the different elements in the array. The "proximity" of individual elements can also affect the overall perception of the scene. "Closure" describes the propensity to fill in small missing pieces of a figure, and perceive it as whole rather than as incomplete in any way.

"Figure/Ground, or Exclusive Allocation" is a phenomenon that has been used to describe the perception of ambiguous figures. In Rubin's "face-to-face/candlestick" reversible figure the simultaneous perception of both possibilities is impossible. Elements within the scene are allocated exclusively to either the figure (the foreground) or the ground (the background) at any one time. However, the figure/ground categorisation is subject to the identity of the object. If transparent figures are imposed on each other the identity of both figures can be perceived simultaneously (Metelli

[2] p.138 Principles of Gestalt Psychology K. Koffka (1935)

1974). The portions of the figures that overlap are necessarily common to the two figures, and as such are not allocated exclusively to the 'figure' or the 'ground'. The Gestalt observation that elements belong either to the ground or to the figure is not true in all cases.

The organisation of any perceptual scene can be described in terms of the principles outlined above, although some of the terms used in Gestalt explanations of how the perceptual system organises a complex scene are ill-defined. It is often difficult to describe why a particular shape is "..the best, simplest and most stable..". Working models of the principles described above have proved difficult to build.

1.4.1 SCENE ANALYSIS

"..the goal of scene analysis is the recovery of separate descriptions of each separate thing in the environment."[1]

In the 1960's research into Artificial Intelligence (A.I.) provided a new approach to the parsing problem. In vision, scene analysis describes how lines and contours in the environment are allocated to different objects. Scene analysis can also be applied to a complex auditory environment.

---

[1] Auditory Scene Analysis A. Bregman (1990)
[1] p.9 Auditory Scene Analysis A.Bregman (1990)

1.4.2 FACTORS INFLUENCING AUDITORY OBJECT FORMATION

Bregman and Campbell (1971) describe a situation in which different parts of the auditory environment are streamed. This phenomenon is described as Primary Auditory Stream Segregation (P.A.S.S.), with each stream indicating a different auditory object, or source. The streaming process is influenced by a number of factors including timbre, rhythm, onset and offset synchrony, spatial location, and spectral and temporal proximity.

1.4.2.a Spectral and temporal proximity

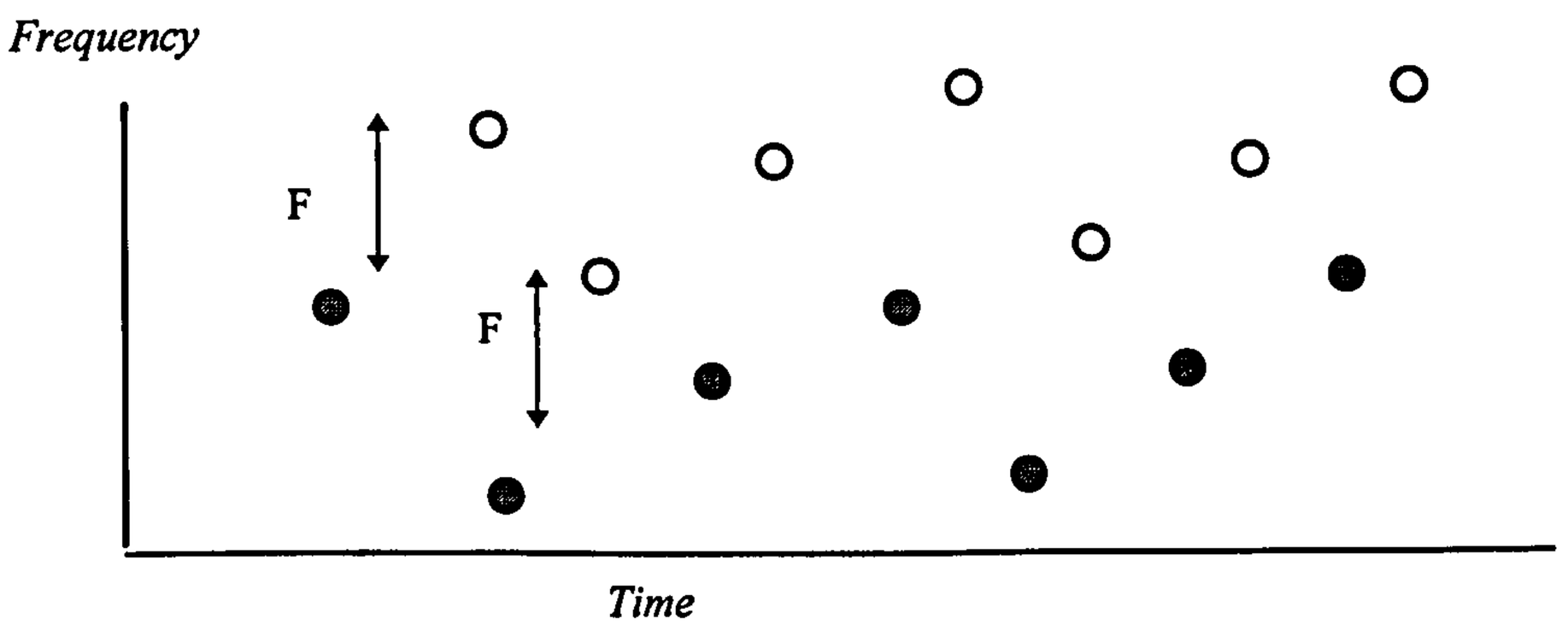


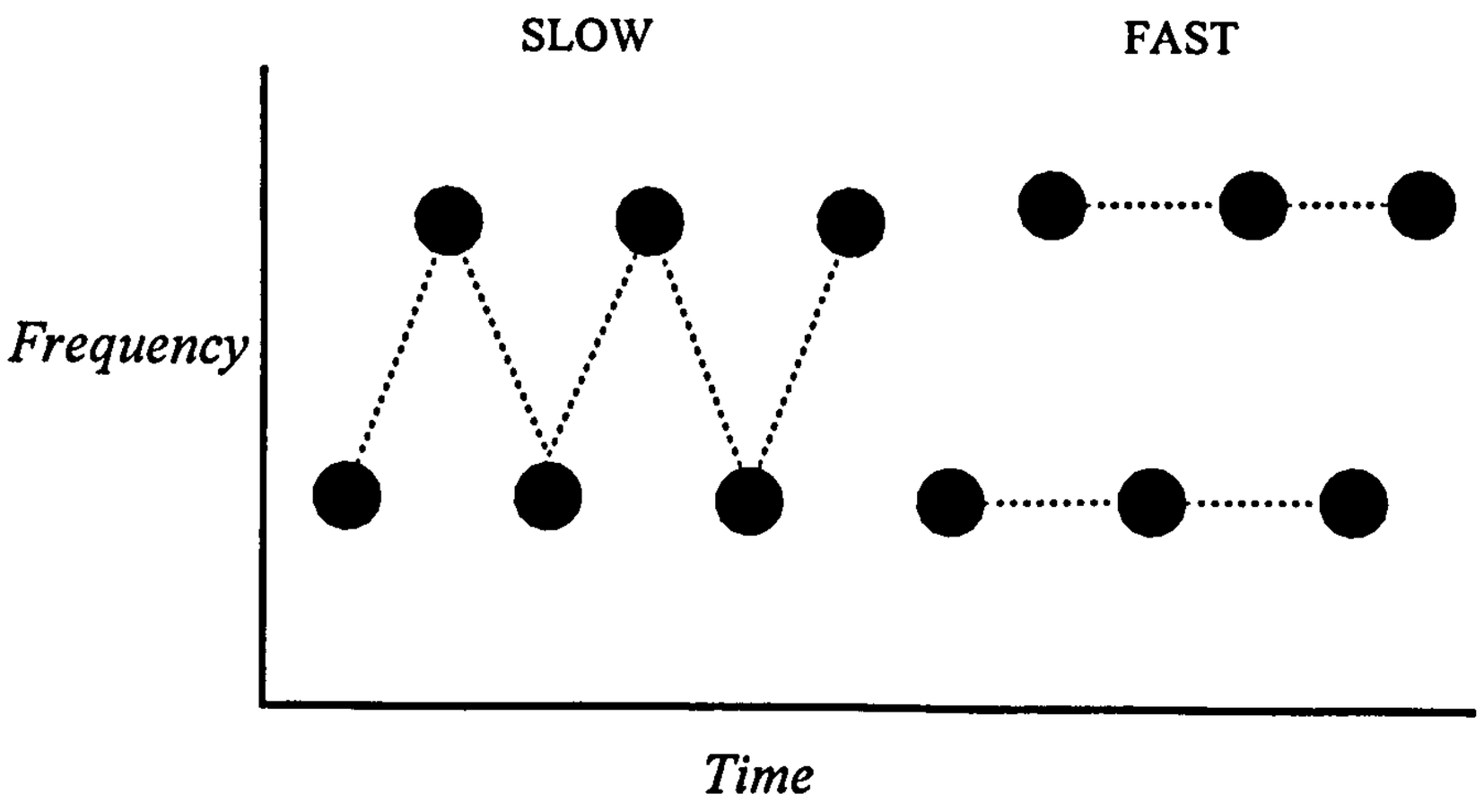
figure 2 (c.f. Bregman and Campbell 1971)

If the sequence of tones in figure 2 is presented slowly, one continuous stream is heard. Fission occurs when the presentation speed is fast enough, so that the single stream splits in two. The higher-frequency white tones are heard in one stream, and the lower-frequency grey tones in another. Fission can also be induced by increasing the frequency separation between contiguous tones (F).

A Gestalt account of the fission process would make reference to the proximity principle, in this case spectral and temporal proximity. Streaming by sequential spectral proximity is made easier by increasing F, or increasing presentation speed. Many theorists in the field regard such organisational streaming as a characteristic of the efficacy of the auditory system, others have suggested elements of the organisational process which could be described as a result of some sort of breakdown of the perceptual mechanism (Van Noorden, 1975; cited in Moore 1989, and Bregman 1990).

Van Noorden (1975) indicates that the temporal splitting of a single fast moving stream of alternating high and low frequency tones into two separate streams determined by frequency (figure 3) is not due to the perceptual mechanism grouping on the basis of frequency similarity, but is instead the result of the over-stretched perceptual system's inability to track a fast moving tone which falls outside a "critical band".

figure 3



PERCEPTUAL STREAM .....

If successive tones in the sequence are separated in frequency sufficiently, they straddle a "temporal coherence boundary" (TCB). Van Noorden (cited in Bregman 1990) showed a trade off between the temporal rate of the tones in the sequence and the frequency separation between successive tones. At high speeds, the separation must be less than 5 semitones for the tones to be fused into one stream, at slower speeds, a higher frequency separation is tolerable.

Two hypotheses were proposed. Van Noorden indicated that the separation may be due to our inability to integrate successive tones into a stream unless they stimulate overlapping populations of hair cells - the overlap hypothesis (Bregman 1990). The hypothesis was based partly on Van Noorden's finding that frequency separation was effective in perceptual segregation, and the knowledge that different frequencies correspond to different positions on the basilar membrane. If the frequency separation between successive tones was large, each tone would be perceived as belonging to a different stream, because, Van Noorden hypothesised, overlapping populations were not stimulated by successive tones in the sequence. Bregman (1990) reports that another reason for the overlap hypothesis was Van Noorden's observation that successive tones in the sequence did not fuse into a single stream if played to different ears, again as a consequence of their not stimulating overlapping hair cell populations. However, Deutsch (1979) showed that successive tones in a melodic sequence, presented to alternate ears are perceived as a single melodic stream if each tone is accompanied by energy (noise) in the non-stimulated

ear. The melodic stream is heard even though successive tones do not stimulate overlapping populations of hair cells. Bregman points out that the hypothesis is further weakened by its inability to address the temporal and frequency trade-off in stream segregation mentioned earlier.

Secondly Bregman (1990) reports a theory proposed by Van Noorden (1975) and Anstis and Saida (1985) suggesting that the problem may lie with 'pitch-change' detectors. Bregman indicates that pitch-change is registered in 'some physiological structure' (p.186) and that the greater the frequency separation between successive tones, the longer the necessary temporal delay between them in order to register pitch-change, although the physiological structure is not identified. Stimuli which demand performance beyond the limits imposed by either or both of these constraints elicit characteristic perceptual streaming of the type indicated.

Bregman and Rudnicky (1975) provide support for the argument that streaming chracterises the efficacy of the perceptual system rather than being a product of an over-stretched perceptual system.

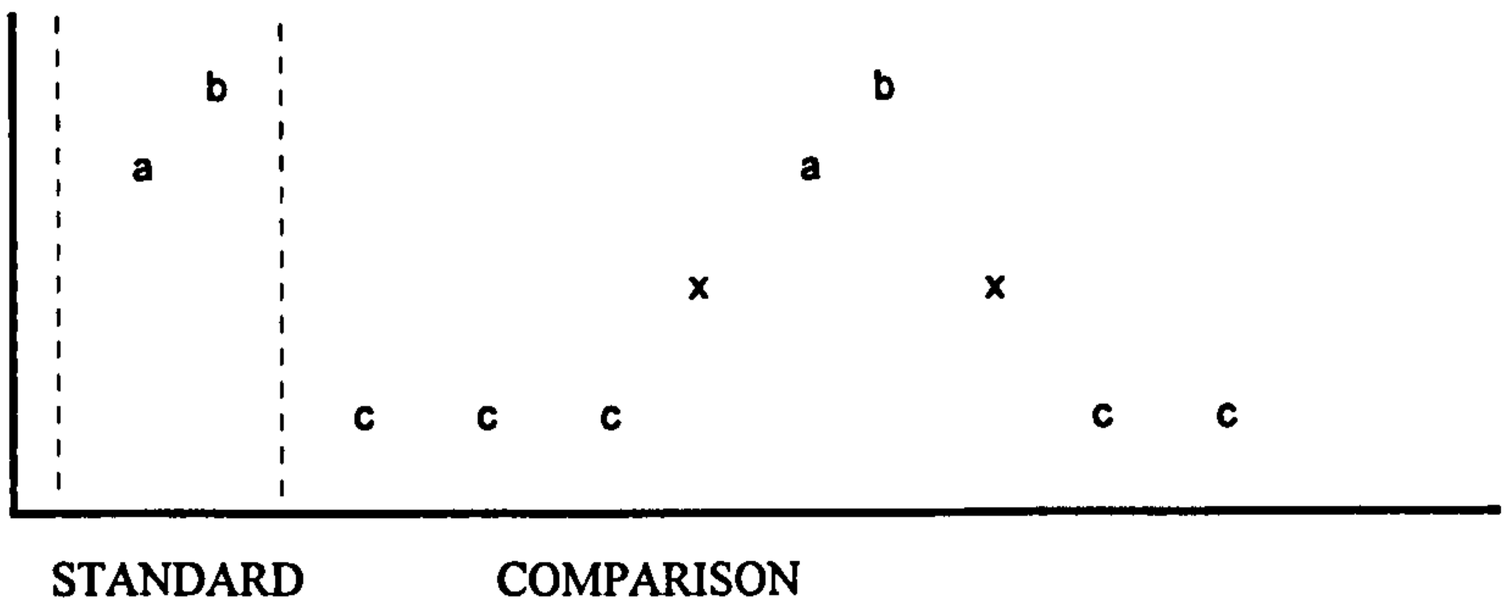


figure 4 Bregman and Rudnicky (1975). 'c' = captor; 'x' = distracter; 'a' and 'b' = targets.

When subjects are presented with the standard a-b, or b-a sequence, a temporal order judgment can be made easily. Embedding the target sequence between two distractor tones makes the judgment considerably harder. Bregman (1990) suggests that this is because the four tones are streamed together and the directional uniqueness of the ab or ba pairing is removed. Introduction of the captor sequence has no effect on subjects' performance on the task if the frequency of the captor tones is considerably lower than that of the distracters. Reducing the frequency separation between the distracters and the captors makes the order judgment easier when the distracter tones are streamed with the captors. The target tones are segregated and their hi-low or low-hi order uniqueness restored making the order judgment easier. A Gestalt account would suggest grouping by frequency proximity but no explanation is offered as to how the streaming is achieved. Bregman and Rudnicky (1975) propose the notion that listeners have an adaptive rejection region - everything falling within the region is streamed together and everything outside is rejected and consequently forms a separate stream. When presented with the embedded

sequence (xabx) the width of the rejection region begins to reduce when the first tone is received. It is suggested that there is insufficient time to reduce the width of the region enough so as to reject the ab pairing. All four tones lie within the region, are streamed together and subjects have difficulty making a temporal order judgment. Adding the captor tones allows more time from the first captor tone to the first tone in the embedded sequence. There is more time to reduce the width of the rejection region. If the captor frequency is considerably lower than that of the distracters, the region will narrow sufficiently to exclude the distracter frequency and the four-tone xabx sequence will be rejected and streamed together. If the captor frequency is close to that of the distracters the boundary frequency of the rejection region will lie between the target pair and the distracters. The target will be rejected and performance on the task improves.

Bregman and Pinker (1978) report a similar finding. They presented a simple tone, X, followed by a complex consisting of two tones, Y and Z. If the frequency of X was far from that of Y listeners reported hearing a simple tone followed by a complex. If the frequency of X and Y was sufficiently close subjects reported hearing two streams, X-Y-X-Y-X-Y and --Z--Z--Z. The tendency to group by frequency proximity removed the percept of a complex tone, splitting it into its tonal components.

Steiger and Bregman (1981) show similar results with sinusoidal glides.

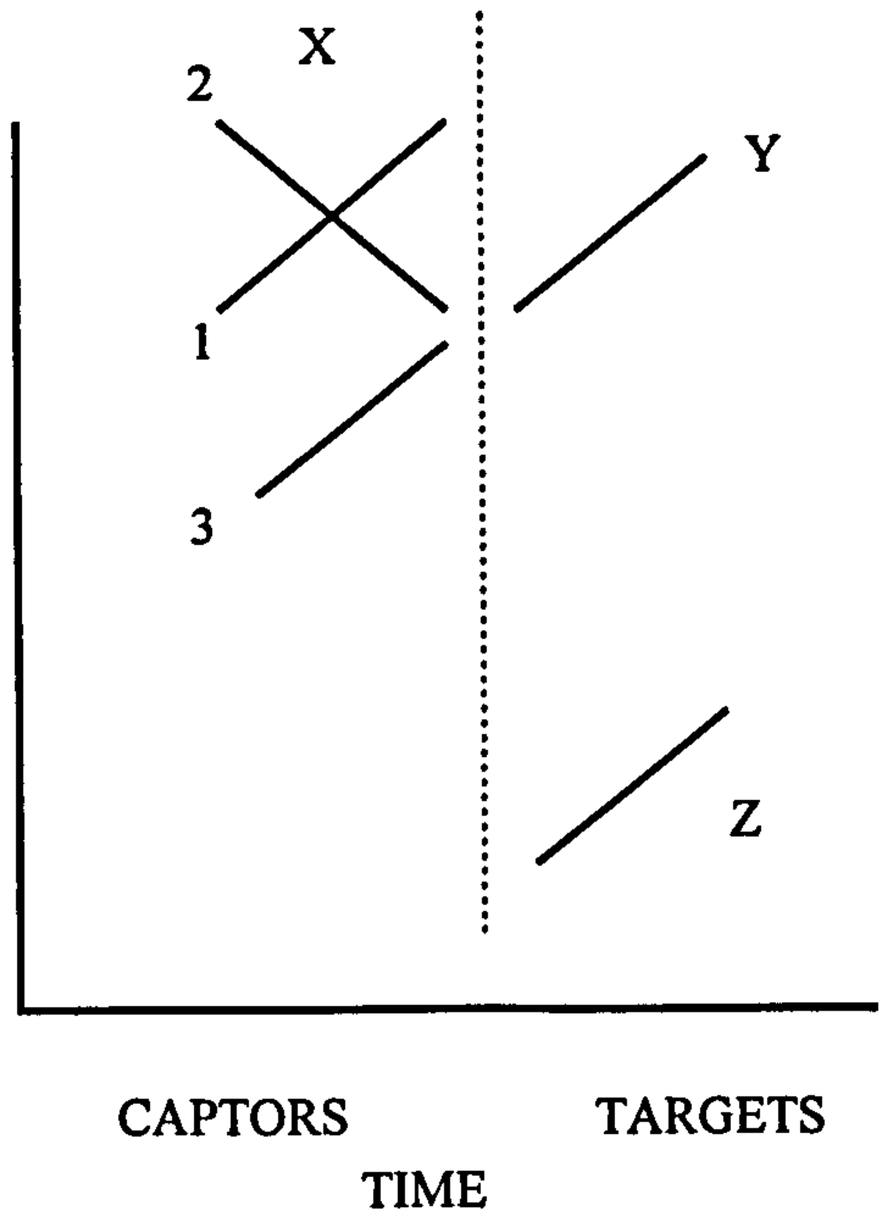


figure 5 Steiger and Bregman (1981)

All glides were 230ms long. When heard in the absence of a captor glide, Y and Z fused and were heard as a single upward rich-sounding sweep. Fission was induced by presenting a captor with an average frequency near to that of the target, in this case X, before the YZ complex. The resulting streams consisted of an X-Y sequence and a single tonal sweep (Z). No effects of trajectory cueing (X3) or proximity of the frequency at the end of the captor and the beginning of the target (X2 and X3) were found.

Although a Gestalt account of the grouping of elements goes some way to identifying the basic variables in the auditory scene analysis process, no satisfactory explanations of how the process works are offered. The force of

the results of Bregman and Rudnicky (1975), Bregman and Pinker (1978) and Steiger and Bregman (1981) is that detailed analysis of the percepts induced by specific stimulus manipulations provides insight into how streaming works. Moreover, the results show that the individual variables in the streaming process (temporal proximity, frequency separation etc.) are not independent. The variables interact, and it is on the basis of this interaction that the perceptual system parses the objects in the auditory perceptual scene.

1.4.2.b Rhythm

Handel, Weaver & Lawson.(83) looked at rhythm as a factor influencing the parsing of an auditory scene and concluded that;

" A rhythm, simple or complex, provides an inherent frame for phenomenal experience: It is not merely a neutral carrier because the rhythm structures that experience."[1]

They describe rhythm as an "intervening variable". The temporal grouping of the elements in the whole scene brings about the perception of rhythmic structure, and it is this structure which provides the basis for stream segregation. The influence of rhythm on the streaming process is complex, and the whole experimental situation must be considered before conclusions about its effect can be made. The authors highlight the importance of considering the experimental context closely before concluding anything about

---

1 p.649 Handel et al. (1983)

the role of rhythm in the streaming process. Temporal and spectral proximity, direction of element movement, whether or not individual elements are of equal chroma are among the many factors which may be relevant to the perception of rhythm.

1.4.2.c Onset and Offset synchrony.

The effect of onset asynchrony on perceptual segregation was shown by Bregman and Pinker (1978). Their experiments showed that segregation of a complex tone into its two tonal components was affected by the frequency of a captor tone presented immediately before the complex. In the same series of experiments Bregman and Pinker found that the perception of two streams was more likely if there was an onset asynchrony between the components of the complex pair.

Rasch (1979) investigated the effect of asynchronous presentation on the detection threshold of a complex tone masked by another complex. If the two stimuli were gated simultaneously the threshold for the target complex was significantly higher than if the target was gated before the masker. If the target was cancelled for part of the time the masker/target complex was present no effect was found - listeners perceived the target tone as continuous and their thresholds remained the same. This could be described as an example of the Gestalt continuation phenomenon, describing the propensity to 'fill in' missing or masked sections of a stimulus assumed to be continuous. Dannenbring and Bregman (1978) presented a sequence comprising a pure

tone alternated rapidly with a three tone complex. The pure tone was set to the frequency of one of the components of the complex in an attempt to stream the two together. They found that segregation could be enhanced by gating the target component of the complex on before the other two components, or gating it off after them, but turning the target on or off within the complex had no effect (c.f. Rasch 1979). A brief preview of the target tone before masker onset (Rasch 1979; Dannenbring and Bregman 1978) or an asynchronous masker-target offset (Dannenbring and Bregman 1978) enhanced streaming, but no effect was found if the target was gated on and off within the masking stimulus. Bregman cites similar evidence from Scheffers (1983) who presented two vowels simultaneously and measured identification thresholds for one of then. He found that if the target vowel was gated on 10ms after the masker, its identification was made easier

Darwin and Ciocca (1992) presented a target complex in one ear, and a harmonic comparison complex in the other ear. The target complex was a harmonic series with one component mistuned and preceding the others by 0 to 300ms. The subjects' task was to adjust the pitch of the comparison complex to match that of the target complex. Results showed that the effect of component mistuning on the perceived pitch of the target complex decreased with increases in the onset asynchrony of the mistuned component relative to the target complex.

Presenting a tone and vowel synchronously can alter the quality of the vowel, measured as the position of the phoneme boundary of an /I/ to /ɛ/ continuum (Darwin 1984). Gating the tone before the vowel reduced its effect on the perceived timbre of the vowel and at 250ms asynchrony it had no effect at all, the conclusion being that an onset asynchrony caused the vowel and tone to stream separately.

1.4.2.d Timbre and brightness.

Bregman (1990) reports an experiment by Wessel (1979) who demonstrated that the brightness of tones can control how a sequence is grouped. "Brightness" is a timbral property of a sound. A brighter sound will have more higher frequency energy than a less bright sound.

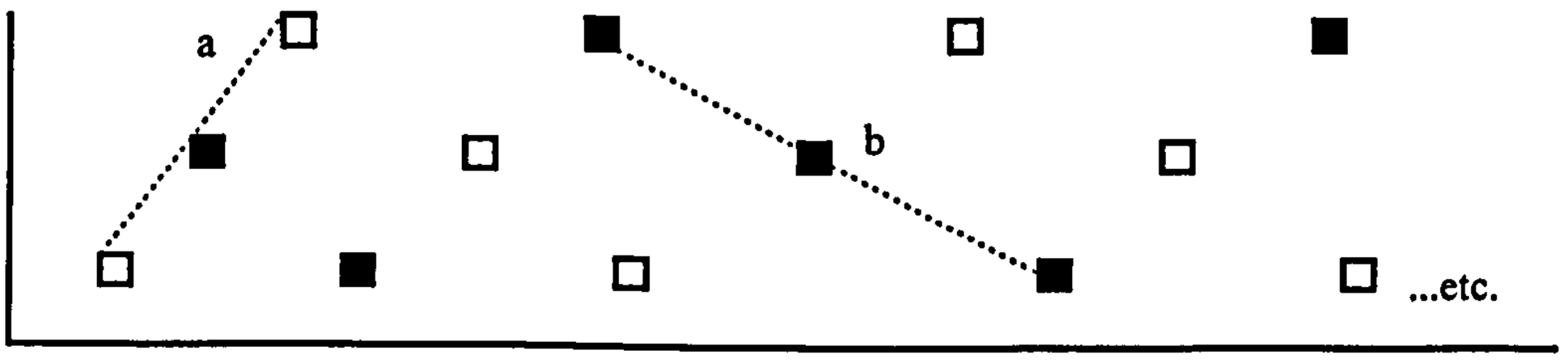


figure 6 Wessel (1979) in Bregman (1990)

If all the tones are of the same brightness repeating triplets (a) are heard. If alternate tones differ in brightness, as shown in figure 6, two slower streams of triplets (b) are perceived, one for each level of brightness.

McNally and Handel (1977) presented sounds with differing timbre in rapid succession and asked subjects to indicate the order of presentation. If sounds of similar timbre were presented next to one another subjects were more likely to report the presentation order correctly. Bregman (1990) points out that this would be expected if the elements were streamed with reference to their timbre. Subjects would report the elements order within their streams, and this would correspond to the order of presentation.

1.4.2.e Spatial Location

If two auditory stimuli come from the same direction they usually belong to the same source. Neurophysiological evidence reviewed earlier shows that a topographic auditory map exists in the Superior Colliculus which could underlie the ability of the perceptual system to identify a common direction. Psychophysical evidence shows that common spatial location is a factor in auditory object formation. Altering the phase, and consequently the lateral position of a target component in an eight tone complex facilitates its perceptual segregation (Kubovy et al 1974). They showed that targeting different components in the complex allowed the perception of a melody. This phenomenon can be explained in terms of binaural masking level differences.

Detection of a binaurally-presented tone masked by noise in each ear is improved by altering its phase in one ear relative to the other. In Kubovy's stimulus, altering the perceived lateral position of the target tone relative to the other tones in the complex effectively increases its detectability and facilitates streaming.

Although all the factors mentioned can be shown to affect streaming in their own right, the experimental situations are somewhat artificial. Handel, Weaver & Lawson (1983) indicate the ambiguity inherent in the available literature. They state that several factors play important roles in the streaming process, but in each case the context must be taken into account, suggesting a possible 'trading relationship' between the organising factors which is likely to account for the enormous flexibility of the perceptual system as a whole.

This thesis is concerned with the lateralisation of audio-visual stimuli and how subjects' lateralisations are influenced by the spatial and temporal correspondence of the auditory and visual elements. Whether the auditory and visual components of the stimulus stream separately or integrate into one stream depends on the Assumption of Unity (AOU). The AOU is itself dependent on a number of factors which will be discussed in the following section, all of which must be considered before the AOU can be calculated. In this respect, the AOU is similar to rhythm, in that the whole experimental situation must be assessed before conclusions about the effects of rhythm or the AOU on the streaming process can be made. All the factors mentioned

here in uni-modal terms are influential in multi-modal grouping, and the principles and structures introduced in this section will be applied in terms of multi-modal integration and streaming.

1.5 MULTI-SENSORY INTEGRATION.

Neurophysiological evidence has shown areas in the brain which are common to two or three different sensory systems, and reactions to bi-modal stimuli are seen even in the very youngest infant. It is clear that information from more than one sensory modality can be combined to form a multi-modal image and it is this combination, or multi-sensory integration, which will be discussed in this section.

Many experiments have investigated the influence of one modality over another by varying the information presented in different modalities. It is generally considered that larger influences of one modality over another can be shown if the subject is convinced that the stimulus presented in different modalities refers to the same multi-modal perceptual event. Ventriloquism is a good example. Manipulating the lips of the mannequin while presenting an auditory stimulus from an unseen or unattended source gives the illusion that it is the mannequin that is talking. The strong assumption that the speech and moving lips refer to the same perceptual event compels the subject to perceive the two physically disparate stimulus sources as common. This particular

assumption is so strong that it overrides the subjects' previous knowledge of inanimate object behaviour.

Radeau and Bertelson (1977) describe this as the assumption of unity (AOU). They indicate that variables which influence the strength of the AOU can be split into two groups, cognitive and structural, with the assumption of unity itself regarded as a cognitive factor. Cognitive factors are those which originate from a familiarity with the type of situation presented, whereas structural factors refer to stimulus-related properties subject to the influence of Gestalt and streaming principles described above, which depend only on the immediate stimulus context. Radeau and Bertelson go on to point out that neither the cognitive and structural categories, nor the individual factors to be discussed, are fully independent of one another, and all must be taken into account in the context of the task in hand. As noted earlier, this is consistent with Handel et al. (1983), who suggested that the context in which the stimuli are presented, and the stimulus properties themselves must be considered before the influence of rhythm in the streaming process is considered.

1.5.1 Cognitive Factors.

The strength of the AOU is only partly dependent on the subjects' awareness of any discrepancy in information provided in the different modalities. Consider the ventriloquist effect. The perceiver is fully aware that the ventriloquist is producing the speech stimulus, although their lips are not

moving correspondingly, and that inanimate objects cannot talk. Despite this they are still convinced that the two modal components belong together, and a AOU is formed. The resilience of the perceptual system to knowledge of inconsistencies in multi-modal relationships is shown in other pairings. Even when subjects are explicitly told that the two modal components to be presented will be discrepant, felt length, slant, texture (Fishkin., Pishkin & Stahl 1975) and limb position (Pick., Warren & Hay 1969; Warren and Pick 1970) can often still be biased by an accompanying visual stimulus.

The 'compellingness' of the perceptual event is also regarded as a cognitive factor affecting the magnitude of any intersensory bias. Compellingness is dependent on a number of factors not least the strength of the AOU, itself a cognitive factor. Welch and Warren (1980) refer to general and specific 'historical' influences on the AOU. Knowing that a noise accompanies an object's collision with a solid surface is gained from general 'history', but specific 'history' allows us to parse footsteps and the accompanying visual stimulus from a complicated multi-modal scene. Amodal characteristics, or those attributes which are common to both sensory modalities also influence the strength of the AOU. A highly compelling perceptual situation would be a case in which perception would be guided by general and specific historical influences arising from past experience, and by amodal characteristics in the sensory inputs, providing evidence for a strong AOU.

The majority of investigations into cross-modal interaction have employed a paradigm wherein the individual modal characteristics of a sensory stimulus presented are made discrepant in some way and subjects' responses to the multi-modal stimulus are monitored. In the example of the ventriloquist the ordinarily amodal spatial location cue is altered. The moving lips and heard words do not emanate from the same point in space. The subjects' response is to relocate the sound to the visual stimulus, indicating a relative dominance, in this case of the visual modality.

1.5.2 Multi-sensory interactions.

Visual dominance over the proprioceptive system is well documented. Shifting the visual component of a task requiring visuo-proprioceptive interaction allows its investigation. Hay, Pick and Ikeda (1965) found a strong influence of vision over felt limb position. The subjects task was to point with an unseen hand to the felt position of their other hand. Part of the target hand was visible to the subject but displaced with prism lenses by up to 16°. Hay et al. found that responses were displaced in the direction of the visual shift. The task is one which has strong general and specific historical components because the correspondence between hand and eye is an everyday occurrence. Subjects reported having no knowledge of any discrepancy in the two sensory inputs suggesting a strong and compelling unitary assumption. Similar findings were reported by Warren and Pick (1970). Perception of the felt length and shape of objects can be altered in a similar way (Fishkin et al. 1975). Pick et al. (1969) showed a proprioceptive influence over audition.

Blindfolded subjects could hear a sound coming from one speaker while their hands were placed on another. The source of the sound was reported as being nearer the position of the felt speaker than its actual source.

The size-weight illusion is another example of a visually-dominated perceptual event. Large containers holding a particular amount of liquid appear heavier than smaller containers holding the same volume. Ellis and Lederman (1993) have shown that both visual and haptic volume cues appear to play a role in the illusion. The strength of the illusion is partly dependent on how the subject picks up the object. Ellis and Lederman showed that a size-weight illusion could arise with haptic or visual cues in isolation. Subjects get sufficient volume information from simply holding or looking at the containers to facilitate the illusion, although a stronger effect was found in the traditional combined visuo-haptic condition. This suggests that cues from the two sensory modalities are combining, not competing. Whereas the influence of the visual component is strong, a role is also played by the haptic sense. In a visually mediated shift of proprioception, Warren and Pick (1970) noted that the influence of the visual component could be manipulated by controlling the amount of the subjects' body which was visible. This suggests that the perceptual system does not rely solely on one or other sensory component but that typically judgments are based on relevant sensory information from all available modalities.

The interaction of two modalities can influence the streaming of a complex perceptual scene. An accompanying visual stimulus can be shown to disambiguate a complex auditory signal. In Cherry's (1953) cocktail party effect, increasing the available perceptual information by providing moving lips enhances subjects' performance relative to a uni-modal condition. In the busy auditory environment of a cocktail party, parsing the auditory scene into streams of individual talkers is made easier if the talkers' mouths can be seen. The additional visual information helps in parsing the now audio-visual scene (c.f. Sumby & Pollack 1954). Everyday experiences confirm that this is the case. Spectacle wearers often indicate that they cannot hear a talker if they are not wearing their glasses and theatre nurses have indicated that if a surgeon's lips are concealed by a mask their instructions are often inaudible. Lip-reading allows subjects to perceive speech correctly under lower signal to noise ratios (MacLeod and Summerfield 1990). When the talkers' lips are visible noise levels can be increased by up to 6dB, and subjects maintain the level of accuracy obtained when just listening (Plomp and Mimpen 1979).

Adding visual information does not always aid in the perception of the scene. When subjects were presented with a video recording of a face repeating '..ga-ga-ga..', with a dubbed synchronised soundtrack of '..ba-ba-ba..', the majority reported hearing '..da-da-da' (McGurk and MacDonald 1976). This phenomenon, the McGurk effect, shows that if a visual stimulus is chosen correctly it can affect perception of the auditory component of the complex. The two types of information combine to yield a perception of the bi-modal

stimulus which is different from the perception of information in each individual modality. This particularly compelling phenomenon is not fully understood. Summerfield's (1987) review indicates that the apparent confusion is not in itself mysterious at all. He suggests that listeners are fully aware of the audio-visual structure of phonemes and when presented with the McGurk and MacDonald stimulus they perceive the phoneme most consistent with the combined evidence from the auditory and visual modalities.

The Fuzzy Logic Model of Perception (FLMP) attempts to formalise this theory (Massaro 1987). It suggests that the visual and auditory inputs are integrated only after independent processing. The audio-visual representation is then compared with a number of prototype memory stores. Which syllable is perceived is dependent on the subject's past experience of audio-visual phoneme perception. More simply;

" X is perceived because it...(the audio-visual stimulus)...looks and sounds like X." (Hearing by Eye. p31)

The factors influencing the combination of heard words and moving lips are essentially the same as the general streaming principles already mentioned. Summerfield (1991) highlights some more specific factors influential in the integration of audio-visual speech.

1.5.3 Audio-visual temporal asynchrony.

Temporal correspondence in the two streams is an important cue to cross-modal correspondence, but the perceptual system is extremely tolerant of temporal asynchronies in audio-visual speech. Measurements of minimal detectable onset asynchrony (auditory leading) range from 80ms (McGrath and Summerfield 1985) to 150ms (Dixon and Spitz 1980). Minimal detectable offset asynchronies (auditory leading) ranged between 140 and 250ms. Common onset of the speech stimulus and an opening of the lips suggests that the two refer to the same perceptual event. Summerfield points out that this alone is not sufficient to indicate to the perceiver that the two streams belong together. Similar dynamics or co-modulation in the auditory and visual stimuli add strong support to an assumption of unity (AOU). The amount of air-flow through the vocal-tract correlates with size of the labial opening and position of the lower jaw and is a determinant of the intensity of the sound. Increasing the lip opening also raises the frequency of the first three formants (House and Stevens 1955 cited by Summerfield 1991). A combination of co-modulation, amodal characteristics and general and specific historical influences leads to a strong and compelling AOU. If the audio-visual asynchrony is increased past these thresholds the unitary assumption is weakened. Radeau and Bertelson (1977) showed that relocation of a sound to a spatially separated visual stimulus (a voice and a film of the talker) was significantly reduced if a 350ms asynchrony was imposed between the two components of the audio-visual stimulus.

1.5.4 Audio-visual Spatial Correspondence; Ventriloquism.

Spatial as well as temporal non-correspondence of the two streams is tolerated within reason. A psuedophone can be used to channel sound which would normally enter one ear to the other by means of a horn-like structure (Young 1928 & Willey at al 1937 cited by Welch and Warren 1980; Kalil and Freedman 1967). Young (1928) concluded that in the absence of vision subjects' localisation of sound was altered by 180°. Provision of the accompanying visual stimulus initiated a return in localisation to the actual source of the sound, an example of the ventriloquist effect. Some leakage in the apparatus allowed the subjects a small amount of direct sound (Willey, Inglis & Pearce 1937). The authors regarded perception of the auditory stimulus in this situation as being suppressed rather than reversed. Provision of information in the visual channel provided more constant spatial information than two conflicting spatial cues in the auditory channel. Suppression of the less consistent auditory channel allowed greater attention to the other modality. Held (1955) carried out similar measurements with an electronic version of the pseudophone apparatus. Subjects reported hearing two images, one in the position indicated by a visual component, the other displaced by 180° (c.f. Willey et al. 1937). Both Willey et al. (1937) and Young (1928) had previously mentioned this dual image phenomenon as 'phantom images' but dismissed it as an effect of the 10% leakage rate of their apparatus. It could be argued that subjects were perceiving two images as a result of inconsistent structural and historical factors. Structurally, the image

should be perceived as displaced by the pseudophone, but historically the source should correspond with the visual stimulus.

Non-speech variants of ventriloquism have been used to investigate manipulations of the compellingness of an audio-visual pairing. Jackson (1953) showed subjects' propensity to indicate the source of a steam whistle as being in the position of a spatially-disparate jet of steam. The sound of a bell was also relocated to a spatially separate light. The spatial separation over which the relocation would take place was dependent on the audio-visual pairing used. The steam-whistle pairing had meaning. Subjects had specific historical experience of steam whistles, but the light-bell pairing was less meaningful. The level of ventriloquism was found to be partly a function of the context of the pairing.

In a similar experiment, the context of the audio-visual pairing was altered by comparing the level of ventriloquism with a voice synchronised with but spatially separated from a talking face and a series of tones synchronised with and spatially separated from a talking face (Thurlow and Jack 1973). The spatially separated voice was relocated to the face over much larger distances than the tone. When the visual stimulus (the face) was replaced with a hand pushing a button the result was reversed, the tone was relocated to the button pushing hand over much larger distances than the voice. The level of ventriloquism experienced with the sound of bongo playing and a video of the drummer was reduced when the visual stimulus was replaced with a video

recording of sound sensitive lights (Radeau and Bertelson 1977). The level of ventriloquism in the two conditions was not significantly different but the difference was large enough to suggest an effect of contextual realism on ventriloquism.

Warren, Welch & McCarthy, (1981) addressed the role of compellingness in the ventriloquist effect. Subjects were required to estimate the magnitude of audio-visual spatial discrepancy in different pairings. Perceived position of the auditory component of the stimulus was indicated on a notional scale, 0 (zero) referring to directly ahead, positive numbers to the right and negative numbers to the left. They were also asked to judge how sure they were that the two components referred to the same perceptual event. Conditions varied in realism, audio-visual synchrony and instructions given. A talking face could be linked synchronously or asynchronously with the corresponding soundtrack or a click train. A piece of tape attached to the VDU screen, covering the mouth of the talker reduced the compellingness of the visual component. Type of instruction was varied by telling subjects that the auditory and visual stimuli to be presented, although spatially separated, would refer to the same event (unitary event instructions), or would not refer to the same event. Results showed an effect of instruction type, level of synchrony and compellingness. Ventriloquism was strongest with unitary event instruction, and synchronous, compelling (voice/mouth) stimuli. They went on to show that audio-visual spatial separation thresholds were smallest for synchronous, compelling stimuli.

1.5.5 Audio-visual interaction in perceptual organisation.

Analogous rules and heuristics to those discussed in terms of the parsing of the visual scene can be applied to the auditory modality - Primary Auditory Stream Segregation (P.A.S.S.). O'Leary and Rhodes (1984) looked into whether organisation of an auditory environment could be influenced by analogous stimuli presented in the visual modality.

Visual stimuli were presented in sequence shown in figure 7. At lower alternation rates one continuously-moving object was perceived, 1-2-3-4-5-6. Increasing the presentation rate resulted in fission into two distinct streams (c.f. Bregman and Campbell 1971). Two apparently moving objects were perceived, 1-3-5 and 2-4-6.

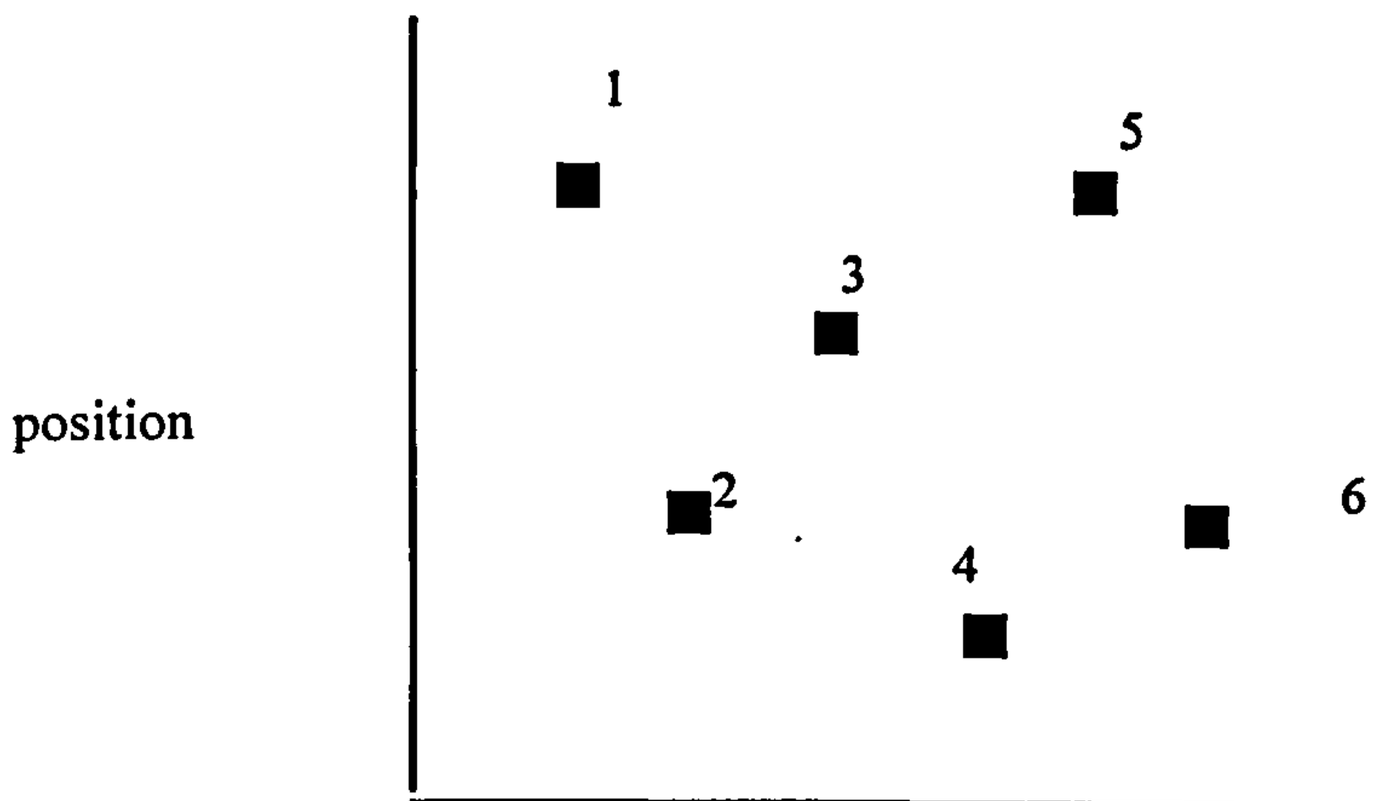


figure 7. O'Leary and Rhodes (1984) page 566.

The auditory stimuli used by O'Leary and Rhodes (1984) were the auditory analogue of the visual stimuli, the pitch of the tone corresponding to object

height in the visual array. If the presentation rate of the auditory component of the audio-visual stimulus was set at a speed sufficient to induce fission, the rate at which the visual stimulus needed to be presented for fission to occur was significantly reduced. The converse was also true. The result suggests that the organisation of one modal component of an audio-visual complex can significantly influence the parsing of the other component.

1.5.6 Audio-visual interaction in the perception of identity duration, and rate.

McGurk and MacDonald (1976) showed how the perception of a syllable can be altered by the simultaneous visual presentation of a different syllable. It is also possible that the identification of a non-speech auditory stimulus may also be significantly influenced by a visual stimulus. A visual influence over the identification of non-speech stimuli is possible with sounds drawn from a Pluck-Bow continuum (Saldaña and Rosenblum 1993). Pluck and Bow sounds and the visual events accompanying them are clearly distinguishable. A pluck sound is a short staccato sound, produced by sharply plucking a string, and characterised by a sharp attack and decay. A bow sound is a smoother sound, with a slower attack and decay produced by drawing a bow over a string. Video recordings of a cellist producing a bowed or plucked sound were synchronised with 450ms sounds taken from a five-point Pluck to Bow continuum. The subjects' task was to rate each audio-visual stimulus on a scale ranging from 0 (zero) for Pluck to 18 for Bow. Subjects were explicitly told to base their judgments only on what they heard, although the accuracy with which subjects followed this instruction is unknown. Results

showed that judgments of all five points on the continuum were significantly more Bow or Pluck like if the corresponding visual stimulus accompanied the sound than if the auditory component was presented in isolation. The results suggest that information in one modality can influence the perception of information in another modality, although it is not possible to say whether the influence of the visual component of the stimulus on the independent variable was a function of an interaction of the auditory and visual components, or whether the influence of the visual component was post-perceptual.

There is evidence to suggest that the relative dominance of the modalities is different for the perception of time and temporal pattern. The results of studies of sensory conflicts in spatial perception typically show a visual dominance of touch over audition. An auditory dominance in the perception of time has been suggested by Walker and Scott (1981) (reported by Welch and Warren), whose subjects held down a key for the perceived duration of a stimulus. When lights and tones of the same duration were presented, lights were perceived as longer than tones; the perceived duration of the audio-visual stimulus was similar to that of a tone presented in isolation but was significantly different from a light presented alone. Estimation of the duration of gaps embedded in light alone, tone alone and tone-light complexes showed similar results, indicating a strong influence of the auditory modality where duration information was presented in two modalities. The results of the experiment discussed in chapter 8 of this thesis suggest that a visual

perceptual lag may have affected the judgements of Walker and Scott's subjects. This point is discussed further in chapter 8.

An auditory dominance in the perception of bi-modally-presented temporal patterns was shown by Welch, DuttonHurt & Warren(1986). 2.5 kHz tones, and LEDs were presented with repetition rates of 4, 6, 8 or 10Hz. Trials were presented uni-modally (visual or auditory only) or bi-modally. The subjects' task was to assign a value corresponding to the rate of the auditory and visual streams. The rate of a reference stream presented directly before the stimulus was assigned the number "2". To indicate the target stream's relative rate as twice that of the reference stream a subject would respond with the number "4". Bi-modal presentations were of low, medium or high levels of rate mismatch. In low mismatch presentations, one stream would be at 4Hz, and the other at 6Hz. In high mismatch presentations one stream would be at 4Hz, the other at 10Hz. Results showed a strong relative influence of the rate of the auditory component in judgments of the rate of the visual component in low, medium and high mismatch presentations.

Both examples of relative auditory dominance (Welch et al. 1986; Walker and Scott 1981) are consistent with the Modality Appropriateness Hypothesis and the Modality Precision Hypothesis, two theories of intersensory bias and dominance.

1.6 THEORIES OF INTERSENSORY INTEGRATION.

1.6.1 Modality Precision Hypothesis (MPH).

The theory suggests that if two discrepant streams of information referring to the same event are presented in different modalities, the modality which experience has shown to be most accurate in registering the nature of that particular event will be relatively dominant. The precise role of past experience in the assessment of modal accuracy is not clear, although it seems fair to assume that the accuracy of the subjects' response provides feedback for the calculation of the accuracy of a judgement based in a particular modality. There is experimental evidence for a modal hierarchy in spatial tasks. Audio-visual spatial tasks (Jackson 1953; Thurlow and Jack 1973; Radeau and Bertelson 1977; Warren et al. 1981) show a bias of spatial information conveyed by the visual component relative to spatial information in the auditory modality, as do visuo-proprioceptive spatial tasks (Hay Pick & Ikeda 1965) Audio-proprioceptive spatial tasks have shown a dominance of proprioception (Pick et al. 1969). These results are consistent with the MPH. It follows that, in localisation tasks at least, the MPH would predict a precision hierarchy with vision at the top, audition at the bottom with proprioception between the two.

Under the MPH, the dominance hierarchy should depend on the dimension being tested. The sensory dominance hierarchy has been shown to be context dependent, differing for temporal and spatial manipulations. For instance, in

investigations described earlier into the multi-modal perception of temporal rate and duration (Welch et al. 1986; Walker and Scott 1981) a relative influence of the auditory modality over visually processed stimuli was shown, a different ordering to the precision hierarchy found with spatial tasks.

However, comparisons of the relative importance of vision and proprioception have produced findings which are contrary to those predicted by the MPH (Fishkin et al. 1975; Power and Graham 1976). A visual bias over a rod's felt orientation is induced by prismatically altering the subjects' vision. The MPH predicts that a reduction in the precision of the visual component should reduce its influence over the haptic component. Fishkin et al. tested this prediction by blurring the prism, thereby weakening the precision of information visually available to the subject, but found no marked effect. Power and Graham found no evidence of an influence of tactual experience on the magnitude of visual influence over felt shape. Experienced and novice potters, two groups with different levels of tactual experience, showed a similar visual bias. The MPH predicts that the different levels of expertise should be reflected in a different level of visual bias arising because of a difference in the relative precision of the two modalities in the two groups. In a similar experiment McDonnell and Duffet (1972) showed that varying haptic precision by wrapping an object in a number of coverings of different thickness had no effect on the visual bias of the haptic sense whereas the MPH would predict an increased visual bias with the reduction of haptic precision.

1.6.2 Modality Appropriateness Hypothesis (MAH).

The MAH is similar in many respects to the MPH. Each sensory modality is assumed to be capable of a number of functions but each has its own task-dependent speciality which best suits its particular information-processing characteristics (Welch and Warren 1980, O'Connor and Hermelin 1972). Exactly what the design characteristics are is not clearly defined. The theory holds that the auditory system is more 'appropriately' designed for making temporal judgments than the visual system, although visually-based temporal judgments are possible. Similarly, auditory localisation is possible but the design of the visual system is more 'appropriate' for spatial judgments. The MAH predicts that, because of this proposed difference in the auditory and visual modalities, when the temporal characteristics of an event are presented in a bimodal context, with both sound and vision carrying temporal information, the auditory system will be relatively dominant. Similarly, when spatial characteristics of an event are presented audio-visually, the visual modality will be relatively dominant. The MAH and MPH predict similar relative dominances in similar contexts but suggest different reasons for the ordering of the dominance hierarchy. The MPH indicates that the ordering is a function of the relative precision of the modalities where as the MAH indicates that the relative precision is itself a function of the differences in the information processing characteristics of the modalities.

O'Connor and Hermelin (1972) suggested that temporal perception might be best facilitated by the auditory modality and spatial perception by the visual

modality. They presented 3 digits successively in a display box through 3 different openings arranged horizontally. This meant that the digits were differentiated temporally and spatially. Similarly, 3 successive digits could be presented auditorily through three different speakers arranged around the subject. Subjects were asked to indicate the 'middle' stimuli. Results showed that when auditory stimuli were presented, subjects indicated the second digit on the majority of trials, i.e. temporal middle. When visual stimuli were presented, subjects indicated the digit which appeared in the central opening of the presentation box, independent of the order in which the digits were presented. When auditory and visual digits were presented simultaneously, subjects indicated spatial middle rather than temporal middle on 99% of trials. The results suggest a strong visual dominance in the task when auditory and visual stimuli were presented simultaneously. O'Connor and Hermelin (1972) suggest that the undefined referent of the word 'middle' was determined by the modality of display. They go on to suggest that their results indicate that the modality of perceptual input induces a temporal or spatial 'set', whereby judgements of the input are either relatively dominated by the auditory modality or the visual modality.

1.6.3 Directed Attention Hypothesis (DAH).

The DAH suggests that any bias in an intermodal relationship is derived from different levels of attention afforded to the modalities. In visually-dominant spatial tasks, the division of attention is in favour of the visual modality. It has been suggested that attention to a visual object is automatically facilitated by muscle activity directing gaze and focus in a particular direction (Posner, Nissen & Klein 1976) Because of this, perceivers have a propensity to attend to the visual stimulus unless characteristics of other modal stimuli give them reason not to. Reisberg (1978) showed that shadowing one of two female voices was improved if the sources of the sounds (two spatially-separated loudspeakers) could be seen. This suggests that auditory attention can be directed by visually attending to the source of the sound. Reisberg, Scheiber & Potemkin (1981) report similar results. Driver (1996) showed that two spatially coincident streams of speech could be disambiguated by providing the face of one of the talkers in another location, but perception of the speech streams was not improved if the face of the talker was in the same spatial position as the sound source. This finding suggests that the relationship between where a person is looking, and therefore the allocation of visual attention, and the source of a corresponding sound, as well as the identity of the visual object is relevant in this particular investigation of audio-visual speech recognition. Shelton and Searle (1980) showed that the sound localisation accuracy was better in the light than in the dark, suggesting that the actual presence of a visual environment can influence the spatial perception of sounds.

A visual bias over a rod's felt orientation (described earlier) is induced by prismatically altering the subjects' vision. The MPH predicts that a reduction in the precision of the visual component should reduce its influence over the haptic component. Fishkin et al. (1975) tested this prediction by blurring the prism but found no marked effect. It could be argued that this reduction in precision would be counteracted by the distribution of more attention to the visual modality. The DAH could then provide an explanation of the result.

1.6.4 A New View of Intersensory Bias (Welch and Warren 1980).

While providing some possible explanations of the processes behind intermodal integration and intersensory bias, the MPH, MAH and DAH have some limitations. A more accurate insight might be provided if the influences of precision, appropriateness and attention distribution were combined, and considered along with the assumption of unity in one model. The model indicates the variables affecting the perceptual result of a discrepant multi-modal situation.

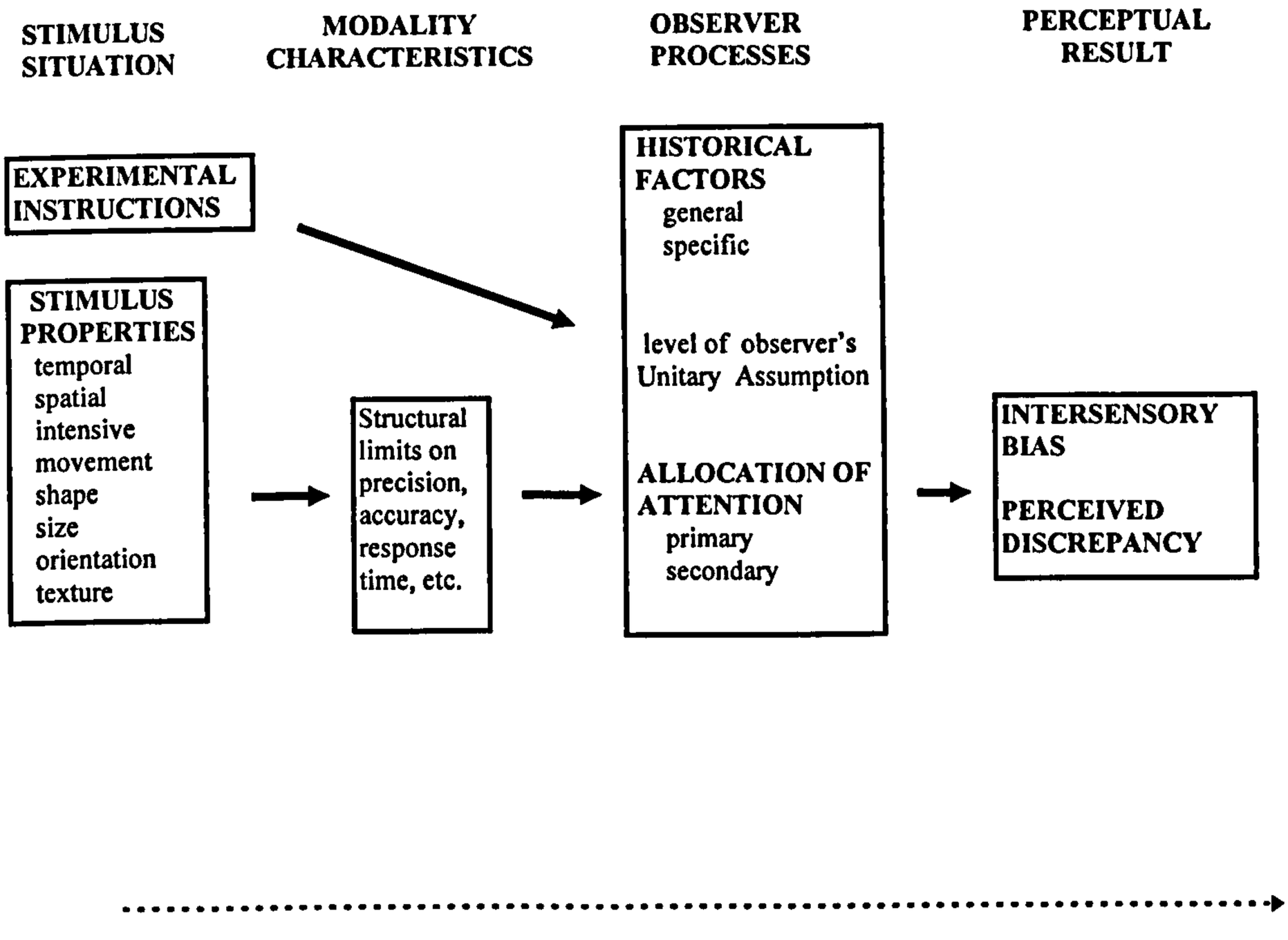


figure 8. Welch and Warren (1980) p.662

·····→ (direction of 'flow')

1. Stimulus Situation.

Properties of the stimulus are received by more that one modality. Any experimental instructions regarding the event influence the assumption of unity - an observer process.

2. Modality Characteristics.

The nature of the receptive system with respect to the stimulus is taken into account. Welch and Warren point out that the way in which the different modalities receive information about an event is different. For instance the shape of an object is received haptically over a relatively long period, as the observer explores the object with their hands. However, visual information

about shape is received relatively quickly, depending on the size of the object. Welch and Warren consider that the relative influences of the modalities must itself be influenced by these 'modality characteristics'

3. Observer Processes.

The subject's general and specific past experience is considered, and an assumption of unity made. The strength of the AOU is a function of experimental instruction and any historical factors, as well as amodal elements - features common to two or more modalities - in the individual modal streams. Stimulus properties are also influential in the AOU formation. Attention is weighted according to which of the sensory systems being stimulated is the most appropriate for the task. The model allows for a secondary adjustment of attention allocation based on task experience, additional instruction, strategy application etc.

4. Perceptual Result.

The result is an intersensory bias if the information in the modalities is discrepant, and the AOU is sufficiently strong. Feedback from the implementation of the perceptual result influences weightings and settings in the assessment of further perceptual scenes.

1.6.5 SUMMARY

- There is neural evidence for the existence of multi-modal centers.
- The Assumption of Unity (AOU) is the assumption that information in the individual modalities refers to the same perceptual event.
- The strength of the AOU may be influenced by cognitive factors such as specific experience with the stimuli, and structural factors such as the temporal and spatial correspondence of the information in the individual modalities.
- The relative dominance of the modalities in a multi-modal task is context-specific. Evidence from audio-visual localisation tasks suggests a relative dominance of the visual modality.
- Three hypotheses of intersensory interaction and bias have been presented; The Modality Appropriateness Hypothesis (MAH), the Modality Precision Hypothesis (MPH) and the Directed Attention Hypothesis (DAH).
- The 'New View of Intersensory Interaction' proposed by Welch and Warren (1980) integrates modality appropriateness, modality precision and attention direction into a model of intersensory interaction and intersensory bias.

1.7 INVESTIGATING AUDIO-VISUAL INTERACTION: PROCEDURAL CONSIDERATIONS.

In situations where multi-modal stimuli are presented, interactions occur between the senses receiving the information. Warren at al (1983) differentiate between two different techniques which can be used to measure intermodal interaction.

Most of the research into intersensory interaction has used a technique described by Warren, McCarthy & Welch (1983) as a 'discrepancy' method. An intermodal discrepancy is imposed within an otherwise corresponding multi-modal stimulus and subjects' judgements analysed to determine the relative influence of each modality on their perception of the event. Hay et al. (1965) investigated the relative dominance of proprioception and vision with a visuo-motor pointing task. Subjects viewed one of their forefingers through prism spectacles which displaced the position of the visual image by 11°, thereby making the otherwise corresponding visual and proprioceptive information spatially discrepant. Subjects were required to point to the felt or seen position of the visually-displaced finger using an unseen finger on their other hand. Results showed a strong influence of vision over proprioception. Felt position was strongly influenced by the seen, displaced position of the target forefinger.

Warren et al (1981) and Radeau & Bertelson (1977) showed similar results with ventriloquism. The auditory and visual components of the stimulus (voices and moving lips) were made spatially discrepant. Results showed that the perceived source of the sound was strongly influenced by the position of the visual component of the stimulus. Radeau and Bertelson describe similar experiments using non-speech, audio-visual stimuli and found corresponding visual influences over audition (c.f. Jackson 1953).

Warren et al (1983) describe another technique for estimating intermodal interaction as a 'non-discrepancy' method. The concept of "tagging" is introduced whereby the relative influences of different modalities can be tracked by investigating, for example, how 'visual' or 'auditory' the results appear to be. The technique identifies response profiles characteristic of each modality using uni-modal stimuli, thereby allowing intermodal interaction assessments to be made using multi-modal stimuli. In order to distinguish between the relative influences of the different modalities in question, the "tag" chosen must have a different value for each modality. Warren et al. compared the relative influences of the auditory and visual modalities assessed using a discrepancy method with their relative influences assessed using a non-discrepancy technique. Variability in localisation of auditory and visual targets in an audio-visual localisation task was used as a "tag". The authors were working under the assumption that variability in localisation of auditory targets was larger than variability in the localisation of visual targets. The relative dominance of the two modalities in the audio-visual task could be

assessed in terms of how similar the variance in responses to audio-visual stimuli was to the variance in responses to auditory or visual stimuli. For the non-discrepancy technique to be viable in the assessment of relative dominance, measurements would need to show that the technique would provide the same assessment of relative modal dominance in non-discrepant stimuli as it would with stimuli with an imposed intermodal discrepancy. The technique allows the possibility of measuring the relative influences of the individual modalities in the perception of natural stimuli as well as artificially generated stimuli with or without an imposed intermodal discrepancy.

In Warren et al's experiment, a male face (presented on a VDU) was paired with his voice, which could be presented in the same position or spatially displaced by 10°. Lateralisation responses to the auditory and visual components of the stimulus were measured in two control conditions. Subjects were required to indicate the perceived position of the visual or auditory component of the stimulus using a rating scale from 0 (straight ahead) to +/- 8 (left/right). They were also required to rate their localisation judgements using a confidence rating scale. After each session in which audio-visual stimuli were presented, subjects made a 'unity' judgement representing their AOU regarding the stimulus, and also a rating of perceived spatial discrepancy between the stimulus components.

Results showed that the magnitude of the standard deviations (SD) was a function of stimulus type and response type. As had been assumed, the

variance in responses to auditory stimuli was larger than the variance in responses to visual stimuli. Variance in judgements of audio-visual stimuli was not significantly different from variance in judgements of uni-modal visual stimuli, independent of the level of audio-visual spatial correspondence. This suggested that judgements of audio-visual stimuli were more like judgements of visual stimuli than judgements of auditory stimuli, and that assessment of relative modal dominance by use of the SD 'tag' was independent of the level of audio-visual spatial mismatch in the stimuli. The authors compared the results with an earlier paper (Warren et al 1981) which asked similar questions using a discrepancy technique. They concluded that:

"The naturally occurring SD index, an index that does not depend on an experimentally induced discrepancy, showed the same pattern of variation with the independent variables that the experimentally induced location discrepancy index did."[1]

1.7.1 EXPERIMENTAL OUTLINE

The experiments described in this thesis used attributes of both discrepancy and non-discrepancy methods. A lateralisation paradigm was employed, in which perceived positions of audio-visual stimuli with spatially corresponding or spatially non-corresponding components were indicated with auditory or visual pointers. Responses were analysed with respect to their mean - a measure of bias in judgements of stimulus position - and their variability - a

[1] Warren, McCarthy and Welch (1983). page 418

measure of response accuracy. Variance in response was used as a 'tag' to indicate the relative influence of the auditory and visual modalities on responses to the audio-visual stimuli.

1.7.2 LATERALISATION

The experiments were concerned with lateralisation of audio-visual stimuli rather than absolute localisation as in the experiments discussed earlier (Warren et al 1981,1983; Radeau and Bertelson 1977; Jackson 1953). The localisation of a sound refers to the judgement of the position of a sound source at any azimuth and elevation in the free field. Lateralisation refers to the percept of a sound presented over headphones appearing to be inside the head (intracranial), in a position on a lateral axis drawn between the ears.

The experiments investigated the relative influences of the auditory and visual modalities as a function of the spatial and/or temporal correspondence of the auditory and visual components of the audio-visual stimuli presented. Although traditionally measured in the context of localisation, lateralisation provides a useful tool with which to investigate this issue.

The apparent intracranial position of a binaurally presented tone can be altered by manipulating the relative phase (Interaural Phase difference, IPD) and/or relative intensity (Interaural Intensity Difference, IID) of the signals presented to each ear.

1.7.2.a Interaural Intensity Difference (IID).

The perceived position of binaurally-presented tones is linearly related to their IID up to approximately ±12dbIID (Watson and Mittler 1965). The range of IID's over which the relationship is linear is partly dependent on the frequency of the tone. For lower frequency tones (200Hz, 500Hz) the relationship between perceived position and IID is linear out as far as ±15dBIID (Yost 1981). At higher frequencies (5kHz) the linear range is reduced to approximately ±9dbIID. That is not to say that tones with much larger IID's are not detectable and informative about a sound's position. At greater IID's, additional intensity discrepancies have a decreasing influence on the perceived position of the sound.

The perceived position of binaurally-presented tones with a particular IID is biased towards the ear at which the intensity is greatest, and is symmetrical about intracranial center (Yost 1981). A binaural tone with a 7dBIID in favour of the left ear will appear somewhere on a lateral axis between the left ear and intracranial center. Presentation of the same tone with an IID of the same magnitude but favoring the right ear will give the impression of a tone on the right side of the head, an equal distance away from the center point.

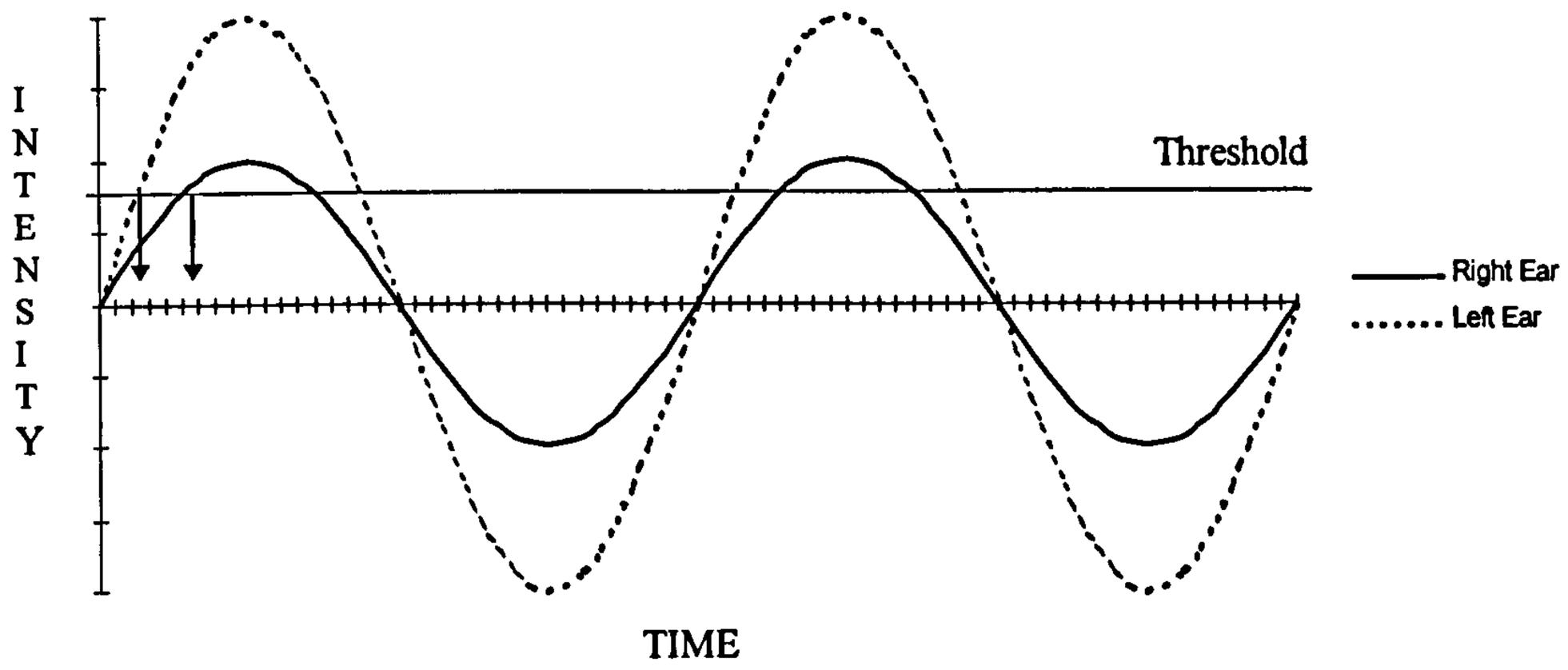
The acuity of the binaural system in discriminating between different IID's varies in a similar way to the minimum audible angle (MAA) in free field localisation. As with the MAA, smaller differences in location are detectable when the sound varies around the central position than when the sounds appear

off to one side (Yost and Hafter 1987). Just noticeable differences in IID are around 0.5-0.75 dBIID at intracranial center rising to approximately 1.4dbIID at around the ±15dBIID position. The thresholds are relatively constant across frequencies ranging between 200Hz and 5kHz except for a marked increase at 1kHz (Yost and Dye 1988).

Grantham (1984) suggests that the raised threshold at 1kHz may be due to a two-component system for lateralisation. The stimuli used by Yost and Dye, and later by Grantham, differed only in interaural intensity; there were no experimentally imposed temporal differences. The lower threshold for tones with frequencies of less than 1kHz must be explained in terms of interaural intensity differences. Grantham assumes that sensitivities to lateralisation cues given by IID and ITD are developed in the free field and are most effective in different frequency regions. At higher frequencies, lateralisation information from the ITD is less accurate than information from the IID in the signal. Grantham (1984) suggests that 1kHz marks the point where the binaural system switches between the two lateralisation cues (IID and ITD), neither operates optimally at 1kHz and spatial acuity is reduced for frequencies in that region. This is consistent with the increase in JND at 1kHz. The author goes on to suggest that at lower frequencies the temporal and intensity comparison systems combine in some way to produce a single temporally-coded lateralisation cue. This would be consistent with Grantham's analysis of an increase in JND at 1kHz as being a symptom of a

two-component system for lateralisation despite the fact that his stimuli did not differ temporally.

Yost and Hafter (1987) show how an intensity difference may be manifest as a temporal difference.



The two sinusoids are matched temporally but differ in intensity. If a threshold of neural stimulation is introduced, the diagram shows how the more intense signal (the broken curve) could evoke an action potential before the less intense signal (the solid curve). The difference in the stimulation times is a function of the difference in intensity between the two signals.

1.7.3 SUMMARY

Two paradigms for investigating intermodal interaction have been identified. Discrepancy techniques describe methods in which a disparity in the modal components of a multi-modal stimulus is experimentally introduced, and relative dominance is measured in terms of mean judgements of the stimulus. Non-discrepancy techniques describe methods in which a "tag" is identified which has different values in each of the modalities under investigation. Responses to stimuli presented uni-modally are compared with responses to the multi-modal stimulus, and conclusions are based on how much like the uni-modal responses the multi-modal responses appear to be. The non-discrepancy technique is useful in that the analysis of relative modal dominance is possible when discrepancies in the stimulus are experimentally imposed as well as when they are not.

Lateralisation of audio-visual stimuli were investigated using a combination of discrepancy and non-discrepancy techniques. It was predicted (c.f. Welch and Warren 1980) that variance in lateralisation of auditory and visual stimuli would differ, providing a useful "tag" with which to track the relative dominance of the auditory and visual modalities in discrepant and non-discrepant audio-visual contexts.

Chapter 2

2.0 PRELIMINARY INVESTIGATION OF AUDIO-VISUAL INTERACTION: LATERAL TRACKING OF UNIMODAL AND BIMODAL STIMULI.

Different techniques have been used to investigate the relative influences of the senses on the perception of multi-modal stimuli (Warren et al. 1983). Discrepancy methods allow the observation of judgements of stimuli with modal components which are in some way discrepant. These techniques have been used extensively in the investigation of ventriloquism with different multi-modal combinations. The relative influence of each modality has been shown to be affected by the subjects' assumption of unity (AOU) regarding the discrepant components of the multi-modal stimulus presented (Radeau and Bertelson 1977; Welch and Warren 1980). A stronger influence of one modality over another is shown if subjects regard the different components of the multi-modal stimulus as referring to the same perceptual event. Radeau and Bertelson (1977) illustrated this point by manipulating the 'realism' of an audio-visual event. Drumming hands combined with synchronous drum beats served as the more realistic stimulus, with drum beats combined with modulated lights as the less realistic audio-visual pairing. They showed that in both cases the perceived position of the auditory component of the stimulus was strongly influenced by the position of the visual component. They also

showed that the components could be spatially separated further in the realistic condition than in the less-realistic condition before the relocation of the auditory component to the visual component was significantly affected. The authors concluded that 'ventriloquism' was stronger for the more realistic condition because of a stronger audio-visual AOU than in the less-realistic condition.

Monitoring the perceptual effect of altering the structural correspondence in the two modal components allows investigation of the relative importance of the different modalities in different experimental situations. The importance of each individual structural factor can also be assessed. This discrepancy technique has been used to look into the visual dominance of proprioception (Hay et al. 1965; Warren and Pick 1970; Fishkin et al. 1975), the proprioceptive influence over audition (Pick et al. 1969), the resilience of the audio-visual system to spatial manipulations in the free-field in speech (Radeau and Bertelson 1977) and non-speech (Jackson 1953), and audio-visual interaction in the perception of identity (Saldaña and Rosenblum 1993; McGurk and MacDonald 1976), rate (Welch et al. 1986) and duration (Walker and Scott 1981).

It was the objective of this experiment to establish whether lateralisations of audio-visual stimuli were more accurate than similar judgements of uni-modal auditory or visual stimuli. The structural factors affecting the audio-visual relationship, specifically the spatial and temporal factors, were manipulated

and the effect of these changes on the relative influence of the auditory and visual modalities was investigated.

Subjects were encouraged to perceive the auditory and visual stimuli as referring to the same perceptual event. Given that the AOU varies as a function of experience with the stimulus (Welch and Warren 1980), the stimulus presentation time was relatively long to encourage subjects to perceive the auditory and visual components as a single audio-visual stimulus. Various amodal and structural factors in the individual modal streams were intended to pre-dispose the AOU and the consequent compellingness of the perceptual event. These included common auditory and visual velocity, common onset and offset and common changes in direction.

The accuracy of the lateralisation judgements was hypothesised to be a function of the extent to which subjects based their judgements on the visual or auditory components of the stimuli (c.f. Warren et al 1983). A metric based on response variance was used to provide a 'tag' with which to track the relative influence of the individual modalities in judgements of audio-visual stimuli.

2.1 Auditory Stimulus

A 12 second 'moving' tone with 50ms rise and decay times at onset and offset was synthesised using the MITSYN software package (Henke 1990). The 1kHz tone was presented binaurally, with headphones at an intensity of 72dB.

A percept of movement was achieved by a linear increase in intensity in one channel with a simultaneous decrease in intensity in the other. It was intended that a tone of constant intensity moving smoothly from one ear to the other would be perceived. The complete stimulus was made up of three, four-second lateral sweeps, moving from the left to the right and back to the left. The stimulus changed direction at the extremes of the sweeps without a delay.

2.2 Visual Stimulus

A red circle, 1cm in diameter with a white central point was animated horizontally across a VDU screen. Each 22cm sweep took four seconds. The full 12 second movement of the stimulus was analogous to that of the auditory stimulus described.

2.3 Audio-visual Stimulus

The auditory and visual stimuli detailed above were presented simultaneously.

2.4 Equipment

Auditory stimuli were presented over Sennheiser HD414 headphones, visual stimuli on a 640x200 VGA display. Auditory stimuli were produced by a Cambridge Electronic Design (CED) 1401 under the control of a Dell system 310 PC.

2.5 Subjects

Twelve subjects took part in the experiment. Pure-tone audiometry showed that all subjects had thresholds within the normal range.

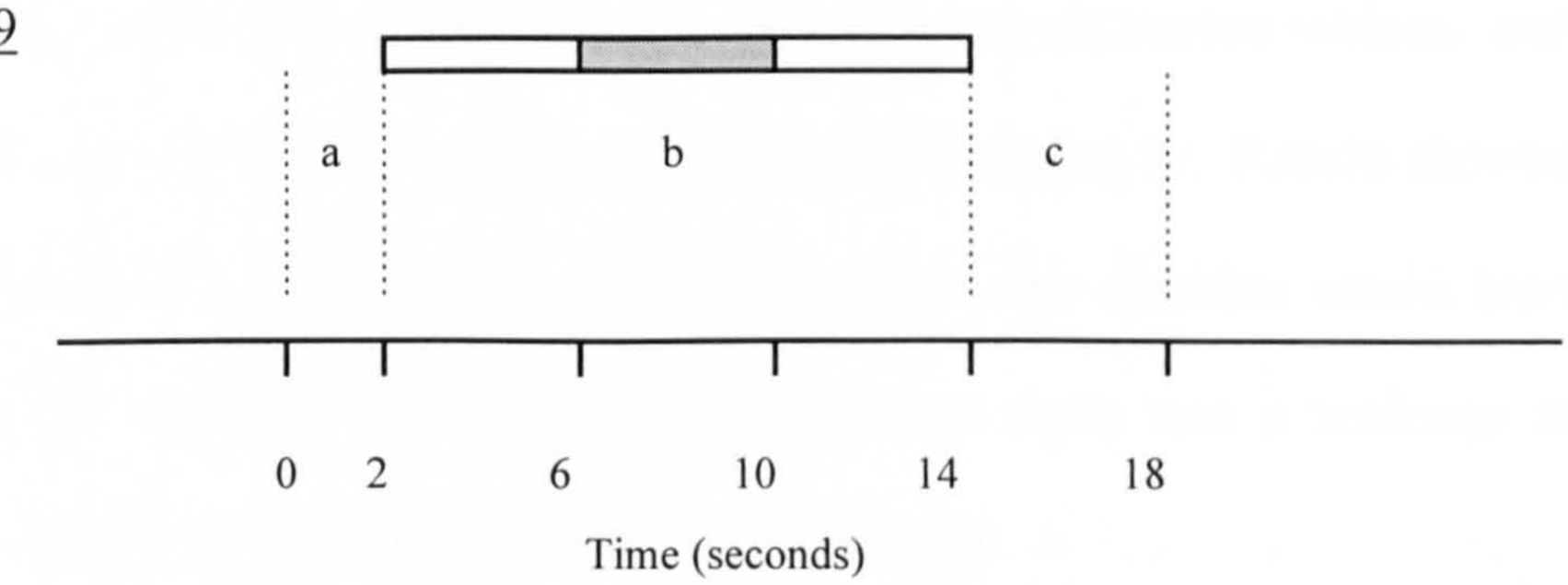
2.6 Procedure

Subjects were seated in a darkened sound-attenuating room with their chins resting on a fixed platform approximately 50cm from the VDU screen. This ensured that visual stimuli were presented at eye level, and that the subjects' distance from the screen was kept constant across trials and conditions.

Trial types were blocked into three different conditions. In the 'auditory' and 'visual' conditions, only auditory or visual stimuli were presented. In the 'combined' condition the auditory and visual stimuli were presented simultaneously. An inter-trial interval of two seconds (a, figure 9) was followed by the stimulus presentation. The stimulus offset was followed by one of sixteen possible delays ranging from 250ms to four seconds in 250ms steps (c). During the delay period a blank screen and no auditory stimuli were presented. At the end of the delay period a visual signal, "RESPOND NOW" was presented. A narrow box indicating the region of the screen traversed by the stimulus, and a visual pointer limited to movement in this region in a random starting position, were then presented on the VDU. The subjects' task was to use a mouse to move the pointer to the position that the visual, auditory or audio-visual image would have reached at the end of the delay period (c)

had it continued back from right to left at the end of the stimulus period (b), and hit a key marked "NEXT" when they were happy with the pointers' position. Subjects were told at the beginning of the experiment that the furthest the image could have traveled during the delay period was full left.

figure 9



(a) INTER-TRIAL INTERVAL	2 seconds
(b) STIMULUS PRESENTATION	12 seconds
(c) VARIABLE DELAY	minimum 250ms, maximum 4 seconds

Subjects received ten practice trials followed by one of the three condition blocks. Each of the sixteen possible delays was presented five times each in random order in each block, making a total of eighty trials per block. Order of condition presentation was fully counterbalanced. Subjects received a different condition on each of three consecutive days.

2.7 Results

The estimated position of the stimulus at the end of the delay period was taken as the horizontal position of the visual pointer in pixels. This allowed analysis of mean responses following each delay, and also errors from the lateral position referring to the point the stimulus would have reached. Mean judgments of stimulus position were averaged across all twelve subjects, and are plotted as a function of delay in milliseconds in figure 10. Results showed a sensitivity of mean judgements to the position the stimulus would have reached at the end of the delay period, although there was a tendency to underestimate the distance traveled in all conditions.

The underestimate of distance traveled increased as a function of delay time. Mean errors from the 'correct' line are plotted in figure 11. A 3-way analysis of variance, with delay duration (16 levels), condition (3 levels), and presentation repetition (5 levels) as factors found a significant main effect of delay duration [$F(15, 165)=18.4$, $p<0.001$] but the mean errors in each of the three conditions were not significantly different from one another [$F(2,22)=0.36$, $p=0.7$]. Presentation repetition was not significant, suggesting that subjects' judgements did not vary significantly as a function of their experience with a particular stimulus [$F(3,33)=1.12$, $p=0.355$).

FIGURE 10: MEAN JUDGEMENT OF POSITION (n=12)

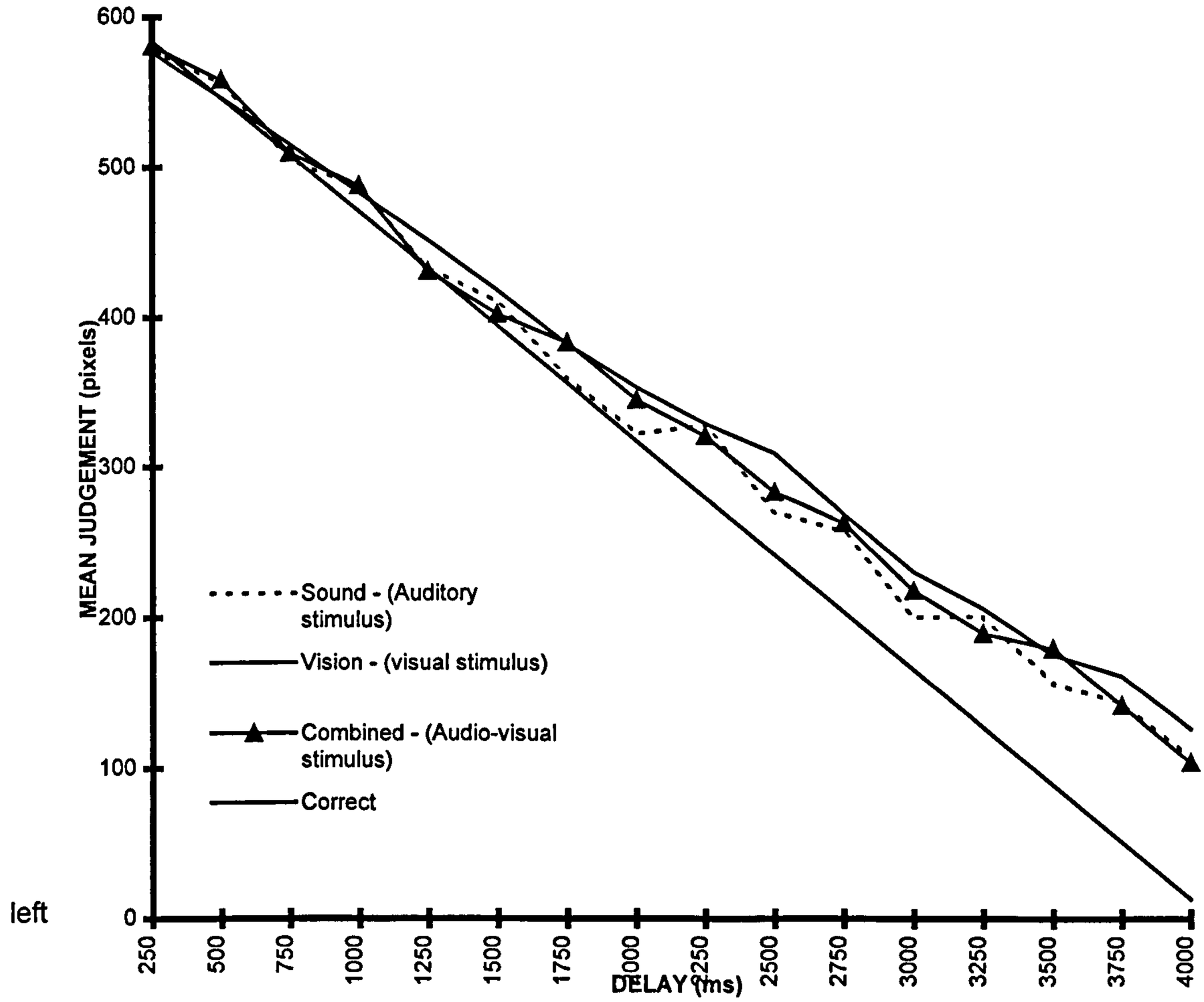
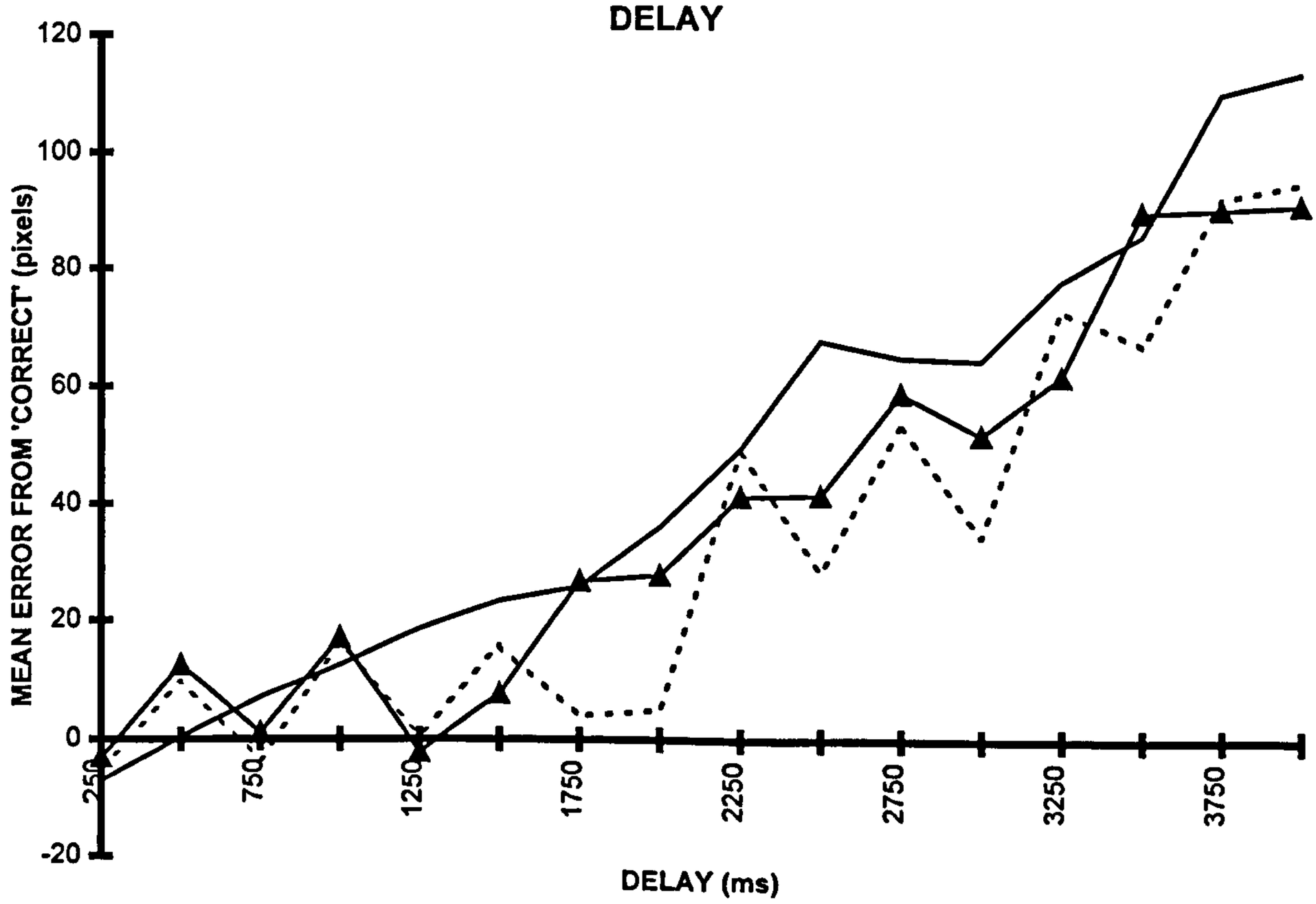


FIGURE 11: MEAN ERRORS FROM 'CORRECT' POSITION OF STIMULUS AVERAGED ACROSS SUBJECTS AS A FUNCTION OF DELAY

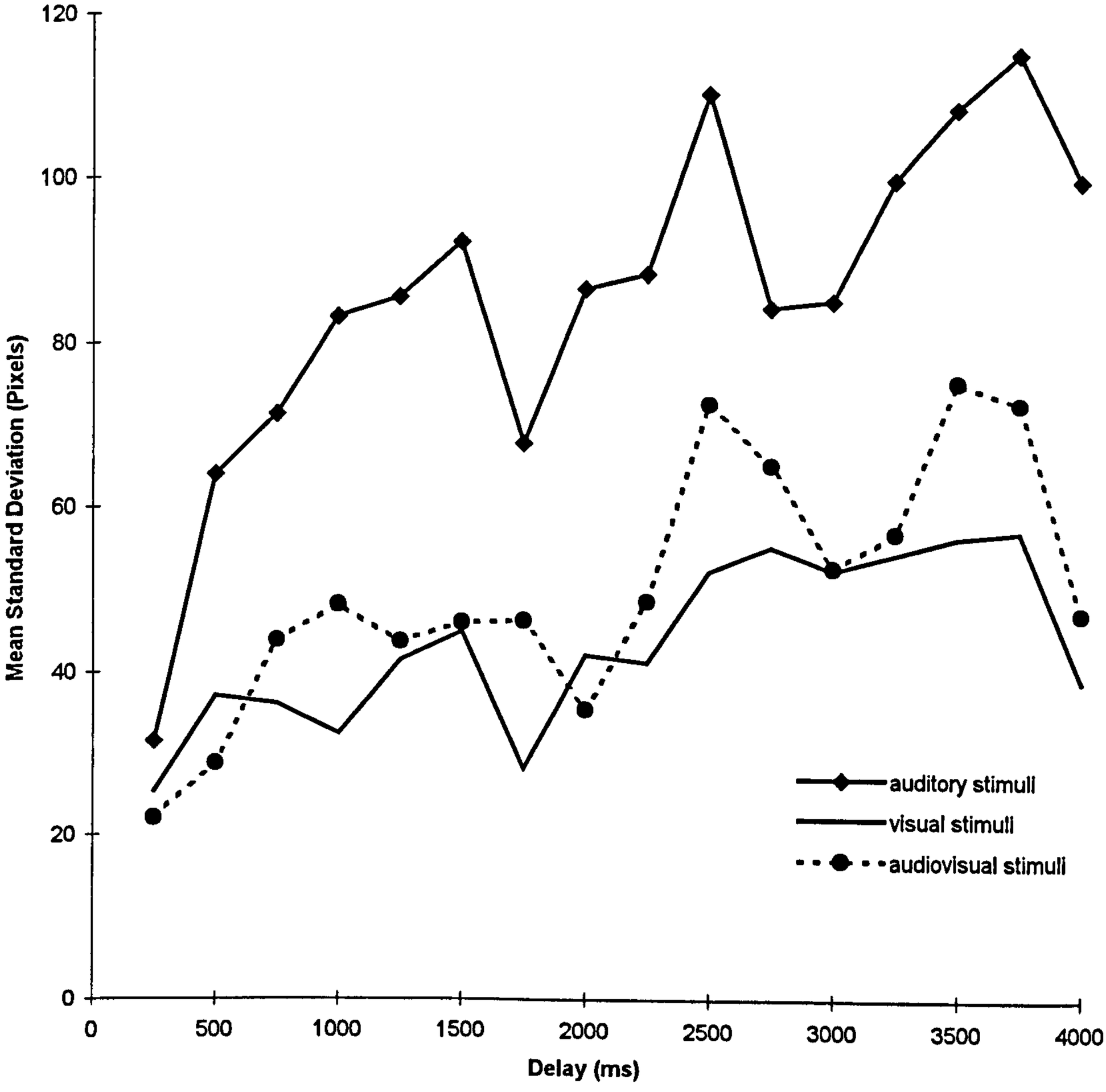


Mean standard deviations in judgements in each condition as a function of delay time, averaged across subjects are shown in figure 12 . Mean standard deviations in judgements of auditory stimuli were greater than mean standard deviations in judgements of visual and audio-visual stimuli. The positive gradient of the functions suggests an increase in the variation of judgements as delay time increased. A two-way analysis of variance with delay duration (16 levels) and condition (3 levels) showed a significant main effect of condition [$F(2,20)=10.25$, $p<0.001$]. Tukey's HSD aposteriori tests - summarised in the table below - showed that the data referring to judgements of auditory stimuli were significantly different from data referring to judgements of visual and audio-visual stimuli. Mean standard deviations in judgements of visual and audio-visual stimuli were not significantly different. A significant main effect of delay was shown [$F(15,150)=7.84$, $p<0.001$]. The interaction between the two factors was not significant [$F(3,300)=0.87$; $p<0.601$].

	Auditory Stimuli (mean = 85.69 pixels)	**Audio-visual Stimuli** (mean = 50.2557)
Visual Stimuli (mean = 43.38 pixels)	Difference = 42.3056 Sig. $p<0.01$	Difference = 6.872 Not Sig.
Audio-visual Stimuli (mean = 50.2557)	Difference = 35.433 Sig. $p<0.01$	

Table 1 - Tukey's HSD comparisons for mean standard deviations in judgements of auditory, visual and audio-visual stimuli. HSD $p<0.05$ = 25.38578; HSD $p<0.01$ = 32.90224

FIGURE 12 Mean Standard Deviations in judgements of auditory, visual and audio-visual stimuli



2.8 Discussion

No significant differences in mean judgements as a function of stimulus type were found (figure 10) consistent with Warren et al. (1981). Deviation from the correct position of the stimulus was found to increase as a function of delay time. The systematic increase in deviation from the 'correct' line suggests that the underestimate of distance traveled was not an edge effect. Compression of judgements at both ends of the stimulus range would have been expected unless the edge effect was unilateral, for example an edge effect associated only with the approach of a 'looming surface'. The data are consistent with a general tendency for the duration of stimuli to be underestimated. Guay (1982) investigated subjects' abilities to reproduce a tone of a given duration. He found that for stimuli longer than one second, subjects consistently underestimated their duration. He went on to show that the underestimate, and variation in the estimate (to be discussed later in this section) both increased as a function of stimulus duration. Schiff and Detwiler (1979) showed similar results in judgements of 'time to collision'. They presented subjects with moving images of plain black disks which appeared to move a short distance towards them. The subjects' task was to estimate the time it would have taken for the object to collide with them if it had continued on the same path at the same velocity. Results showed a linear relationship between the actual time it would have taken for the object to collide and the judged time to collision. However, times to collision were consistently underestimated, and the underestimate and the standard deviation of the underestimate increased as a function of actual time to collision. These studies

are consistent with the assessment of the mean underestimate of distance traveled shown in figure 10 as being a systematic underestimation of the delay time. The longer the time to be estimated, the larger the error (c.f. Guoy 1982).

Further similarities in these data and those of Guoy (1982) and Schiff and Detwiler (1979) are found in the standard deviations of the mean judgements. Guoy (1982) showed that mean standard deviations increased as a function of the time to be estimated, and Schiff and Detwiler (1979) showed similar results. A corresponding increase in standard deviation with delay time was found in the data presented here (figure 12).

It had been the intention that the task would require the observer to use both temporal and spatial information. However, the systematic increase in the underestimation of delay, and the increase in standard deviation as a function of delay time both suggest that subjects may have relied primarily on temporal information. It was possible for the response be based only on the temporal aspects of the stimulus, and not the spatial aspects. In order to complete the task, subjects had to estimate the time between the offset of the stimulus and the onset of the response signal, and convert this time period to distance traveled, based on a stimulus moving at a constant velocity. The time period to be estimated was independent of the stimulus type presented, and the similarities in mean data across stimulus type (figures 10 and 11) are consistent with this independence. Since the task required subjects to combine

knowledge of the stimulus' velocity with their estimation of delay time, it may be that greater attention to the spatio-temporal aspects of the stimulus could have been encouraged by varying the velocity of the stimulus on a trial to trial basis, and that as a result differences in mean position may have emerged. However Schiff and Detwiler (1979) investigated judgements of time to collision as a function of stimulus velocity and found that judgements were not affected.

It has already been noted that the time period to be estimated was independent of the stimulus type presented. It was possible that this independence reflected a ceiling in performance which itself was independent of stimulus type. Differences between subjects' judgements on repetitions of each stimulus type were not found to be a significant factor in the analysis of variance, suggesting that performance on the task did not improve as a function of experience with the stimulus. It follows that if the results reflect a ceiling in mean performance, then the ceiling was reached immediately and as such suggests that subjects found the task too easy.

The results did show that SDs differed as a function of stimulus type (figure 12), as hypothesised. SDs in judgements of auditory stimuli were significantly larger than SDs in judgements of visual stimuli and audio-visual stimuli (c.f. Warren et al 1981), although SDs in judgements of visual stimuli and audio-visual stimuli did not differ significantly. The results provide support for the notion that SDs can provide a useful "tag" with which to track the relative

importance placed on the different modalities when subjects are presented with multi-modal stimuli. In these data, the SD "tag" suggests that information in the visual modality was dominant when audio-visual stimuli were presented, since SDs in judgements of audio-visual stimuli were similar to SDs in judgements of uni-modal visual stimuli.

Figure 10 shows that mean judgements were independent of stimulus type (c.f. Warren et al. 1982). It was suggested earlier that the MAH and MPH (Chapter 1) would predict that optimal information for the spatial demands of the task are provided by the visual modality, and optimal information for the temporal demands of the task are provided by the auditory modality. It follows that the presentation of the audio-visual stimulus should have provided subjects with optimal conditions on which to base their judgements in the spatially and temporally demanding task. If this were the case, judgements of audio-visual stimuli would have been expected to be more accurate than judgements of auditory or visual stimuli. Figure 12 shows that judgements of audio-visual stimuli were consistently more accurate than judgements of auditory stimuli, although no real advantage of audio-visual stimuli over visual stimuli was shown. It is possible that performance may have reached ceiling, and that the comparative ease with which subjects completed the task may have masked any relative advantage of their having been presented with audio-visual stimuli rather than auditory or visual stimuli. The possibility that performance reached ceiling was considered in the design of later experiments.

Welch and Warren (1980) have suggested that response modality may be a factor in an investigation of intermodal integration. If this is the case, the task described here favored the visual modality. The end of the delay period was indicated visually, and responses were made with a visual pointer. This served to direct attention to the visual modality at a crucial time in the task. The Directed Attention Hypothesis (DAH) suggests that the influence of one modality over another is increased by just such an attentional cue. Any differences in perception arising from the type of stimulus presented could have been hidden by this direction of attention at the end of the stimulus to the visual modality.

2.9 Conclusions

It was the objective of this experiment to establish whether lateralisations of audio-visual stimuli were more accurate than similar judgements of uni-modal auditory or visual stimuli. The SD tag has been shown to be useful in distinguishing between the relative dominance of the modalities in a multi-modal context (c.f. Warren et al 1983). The results suggested an advantage of audio-visual stimuli over auditory stimuli, although the accuracy of judgements of audio-visual stimuli was no greater than the accuracy in judgements of visual stimuli. Suggestions have been made regarding possible ceiling effects and a possible temporal equivalence of the auditory, visual and audio-visual stimuli with regard to the completion of the task.

2.10 Implications

Further investigation of spatial, temporal and spatio-temporal audio-visual factors have taken the weaknesses of this experiment into account. The response modality has been considered as a possible factor in the lateralisation of auditory, visual and audio-visual stimuli. If it is the intention to investigate both spatial and temporal factors, the experiment should be designed to ensure that subjects must use both spatial and temporal information for the successful completion of the task. The relative influence of the spatial and temporal factors in judgements of auditory and visual stimuli must be assessed before their interaction in a spatially and temporally demanding task is investigated.

Chapter 3

3.0 GENERAL PROCEDURE AND STIMULI

3.1 BACKGROUND

The task in the preliminary investigation (chapter 2) had both spatial and temporal aspects. It was suggested that the temporal characteristics of the stimuli were independent of whether the stimulus was auditory, visual or audio-visual, and that similarities in mean judgements may have been a function of this independence. It was concluded that assessments of the relative influences of audio-visual spatial correspondence and temporal correspondence in the audio-visual stimuli should be made before investigations into their interaction in a stimulus with spatially and temporally non-correspondent components were made. The modality of response may be a factor in whether the task is considered auditorily or visually based. The Directed Attention Hypothesis (DAH - chapter 1) suggests that the modality to which attention is directed will be relatively dominant in a task. It may be that attention direction is afforded by the modality in which responses are required. For this reason, judgements of stimuli with auditory and visual pointers need to be compared.

The experiments to be described investigated the relative influences of the auditory and visual modalities using a lateralisation procedure, in which the

subjects' task was to indicate the perceived lateral position of a stimulus using an auditory or visual pointer. The spatial and temporal correspondence of the modalities was manipulated in an audio-visual context, and the relative influence of each manipulation on the resulting lateralisation judgement was assessed.

3.1.1 Lateralisation

The lateral position of an auditory stimulus is given primarily by a comparison of the signals in the two ears. Yost (1981) presented subjects with stimuli which ranged in frequency between 200Hz and 4kHz. The IID of the stimuli ranged between ±18dB IID, that is 18dB in favour of the left or right ear. The task was to listen to the target stimulus, presented over head-phones, and indicate its position by moving a slide potentiometer positioned between the ears of a head silhouette, to the position they felt best matched that of the target stimulus. Results showed a linear relationship between IID and perceived position for all of the frequencies tested. The range of IID's over which the relationship was linear was dependent on the frequency of the stimulus, but broadly speaking, the linearity began to break down at approximately ±12dB IID.

3.2 RESPONSE METHODS

Both modalities provide the subject with localisation information. It follows that it should be possible to make a lateralisation judgement of an auditory

stimulus using a visual pointer. Similarly, judgements of a visual stimulus can be made with an auditory pointer.

Lateralisation judgements made with auditory pointers have been shown to be accurate and reliable (Bernstein and Trahiotis 1985; Schiano, Trahiotis & Bernstein 1986). In both studies, subjects were required to indicate the perceived lateral position of an auditory stimulus using an auditory pointer. The IID, and therefore the lateral position of a 200Hz band-pass noise centered on 500 Hz could be varied using a single-turn potentiometer. The experimenters settled on this particular format for the pointer on the basis of a number of preliminary measurements which indicated that it gave a more punctate image than a 500 Hz pure tone.

Assessments made by this experimenter have indicated that a similar level of accuracy was shown with a 500Hz pointer, a 2kHz pointer and the pointer detailed above. It was considered that the 2kHz pointer gave the smaller, more punctate image. The higher frequency pointer was also chosen to allow clear distinction between the pointer and the stimuli to be lateralised (tones of 250Hz).

Adjustments of the pointers provided by Bernstein and Trahiotis (1984) and Schiano, Trahiotis & Bernstein (1986) were made by turning the knob of a potentiometer. The experimenters were aware of the potential problem of subjects using the position of the knob as an indicator of pointer position.

They attempted to counter this by randomising the position of the knob that referred to 0 dBIID (intracranial center) and introducing a random starting position in each trial. The positions of the pointers used in pilot measurements made in connection with the experiments to be discussed here were controlled with a mouse. Pointers began in a random starting position and were adjusted by moving the mouse to the left or right as required. The pointer tone was a repeating cycle of 50ms tones with 100ms silent intervals during which the IID of the tone could be reset relative to the position of the mouse. It was suggested that the absolute position of the mouse could be used as a guide in positioning the pointer. For this reason a large track-ball was introduced in place of the mouse for the experimental sessions.

In summary, the **auditory pointer** was a 2kHz, 50ms tone repeating every 100ms, at an intensity of 72dBSPL, beginning with a randomly-assigned IID. The IID of the tone, varied with a track-ball, was updated during the 100ms intervals. The IID could be varied within a ±18dB range

The **visual pointer** was analogous to the response method used by Yost (1981). His task required subjects to listen to the target stimulus and indicate its position by moving a slide potentiometer positioned between the ears of a head silhouette to the position they felt best matched the intracranial position of the target stimulus. In the experiments to be discussed here, visual responses were made by moving a small arrow between the ears of a blue head silhouette on a black background presented on 640x200 VGA display,

(Appendix II). The pointer was moved with a track-ball and confined to a 24cm axis drawn between the ears of the head outline. Subjects responded by pressing a key when they had positioned the pointer.

In a number of pilot studies, the duration and number of repetitions of an auditory stimulus were investigated in terms of their effect on subjects' lateralisation responses. Four subjects provided data, all of whom had pure-tone thresholds within the normal range.

3.3 STIMULI

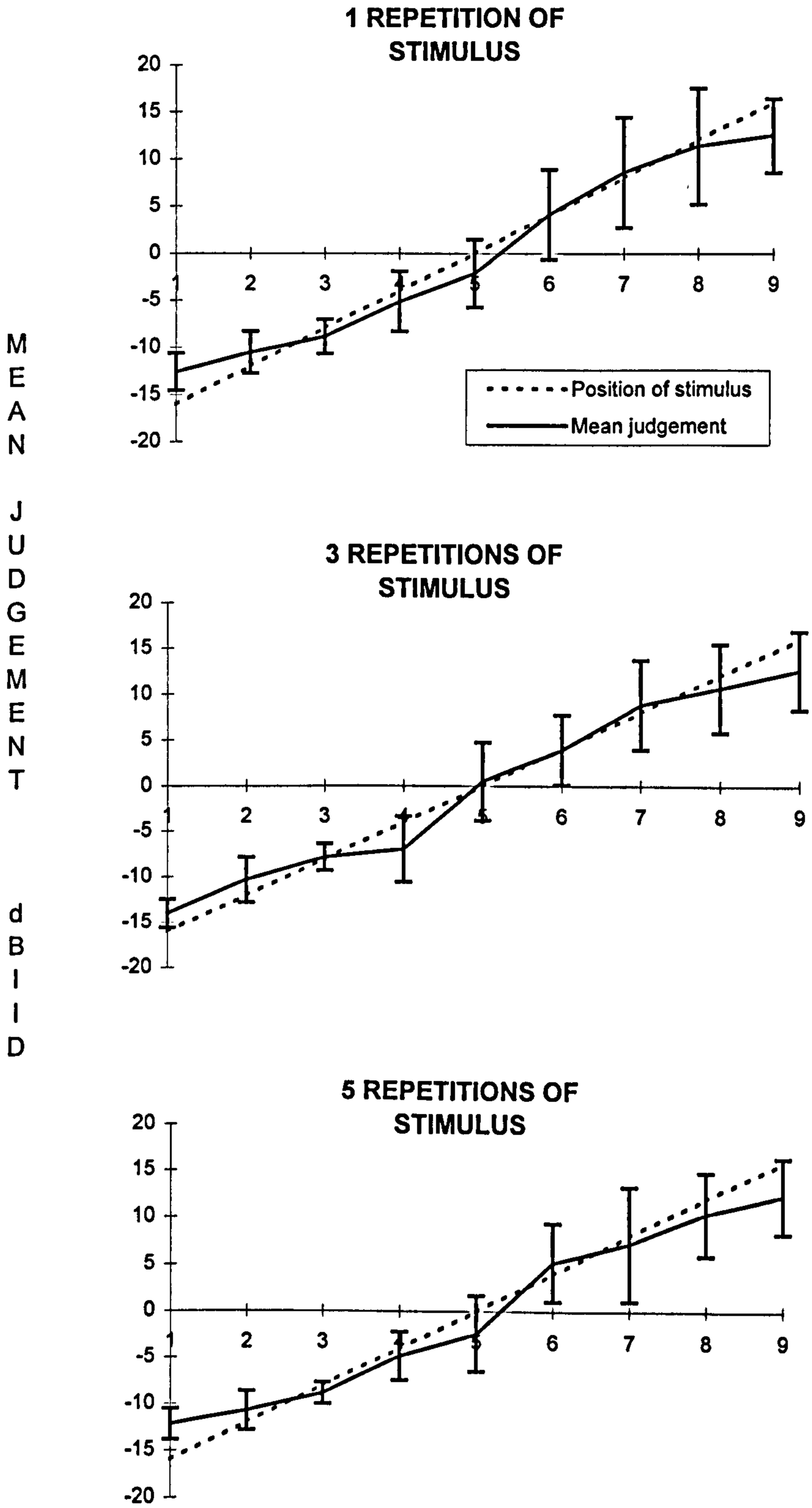
3.3.1 Auditory stimuli

In any one block, stimuli of only one configuration (duration and number of repetitions) were presented in one of nine possible lateral positions given by the IID of the stimulus in the range ±16 dbIID. Stimuli of 25ms, 50ms and 100ms were presented at a frequency of 250Hz. Stimuli could have one, two or three repetitions, with 50ms intervals between each repetition.

3.3.2 Procedure

Stimuli and pointers were presented over headphones and subjects were instructed to move the pointer (presented after 500ms after stimulus offset) so that its position matched that of the stimulus presented. In all trials subjects responded with the auditory pointer described earlier. A typical set of results from the pilot series is shown in figure 13. In the example shown, subjects were presented with 250Hz tones of 100ms in duration.

FIGURE 13: RESULTS OF PILOT JUDGEMENTS OF 100ms, 250Hz TONES



A three-way Analysis of variance with stimulus repetition (3 levels), stimulus duration (3 levels) and stimulus position (9 levels) showed no significant main effects of the number of stimulus repetitions [($F(2,8)= 0.05$, $p<0.949$] or stimulus duration [$F(2,8)= 0.89$, $p< 0.447$]. Stimulus position was shown to be significant [$F(8,32)=69.10$, $p<0.000$]. The two and three-way interactions were not significant. A linear relationship was observed between the IID of the stimulus presented and the perceived position for stimuli with IID's in the range ±12dB. The apparent non-linearity at extreme IID's is consistent with an 'edge' effect. 'Edge' effects are characterised by systematic response errors at stimulus maxima and minima. In these data the edge effect could also be a deviation from the pointer maxima and minima rather than the stimulus maxima and minima. The data suggest a compressive non-linearity between perceived lateral position and stimulus IID for IIDs greater than approximately ±12dB, the 'edge' compressing the responses into a narrower range. In these pilot experiments stimuli of up to ± 16dBIID were presented, and the pointer ranged only as far as ±18dBIID. Potential 'edge effects' arising from a compression of the pointer range were minimised in the experiments to be discussed by limiting the stimulus range relative to the pointer range, so that tones were presented in one of seven possible lateral positions corresponding to IIDs in the range ±12dBIID. The range was chosen so as to be in the linear part of the mean response functions in the pilot experiments. Since no differences were found between subjects' abilities to use the auditory pointer to indicate lateral position in any of the stimulus configurations tested, 1 repetition of 50ms, 250Hz tones was chosen for the auditory stimulus.

3.3.3 Visual stimuli

Visual Stimuli were light-coloured, 2x2-pixel (approximately 2mm in diameter) spots, subtending 13.75 minutes of arc, presented for 50ms in one of seven possible positions equidistant on a lateral, 22cm axis drawn between the ears of the black head silhouette on a blue background. Pilot studies showed that the visual stimuli of the durations used may have gone unnoticed. For this reason, a small white spot cued the position of any visual component at the start of each trial (see figure 14). Pilot studies showed that the multiple repetition of auditory or visual stimuli did not influence the subjects' judgements of lateral position, and as such a visual cue would not afford any bias to the lateralisations of visual stimuli. As above, the range covered by the visual pointer was greater than the range over which stimuli varied, to minimise potential edge effects which may affect judgements of more extreme visual positions.

3.4 EQUIPMENT

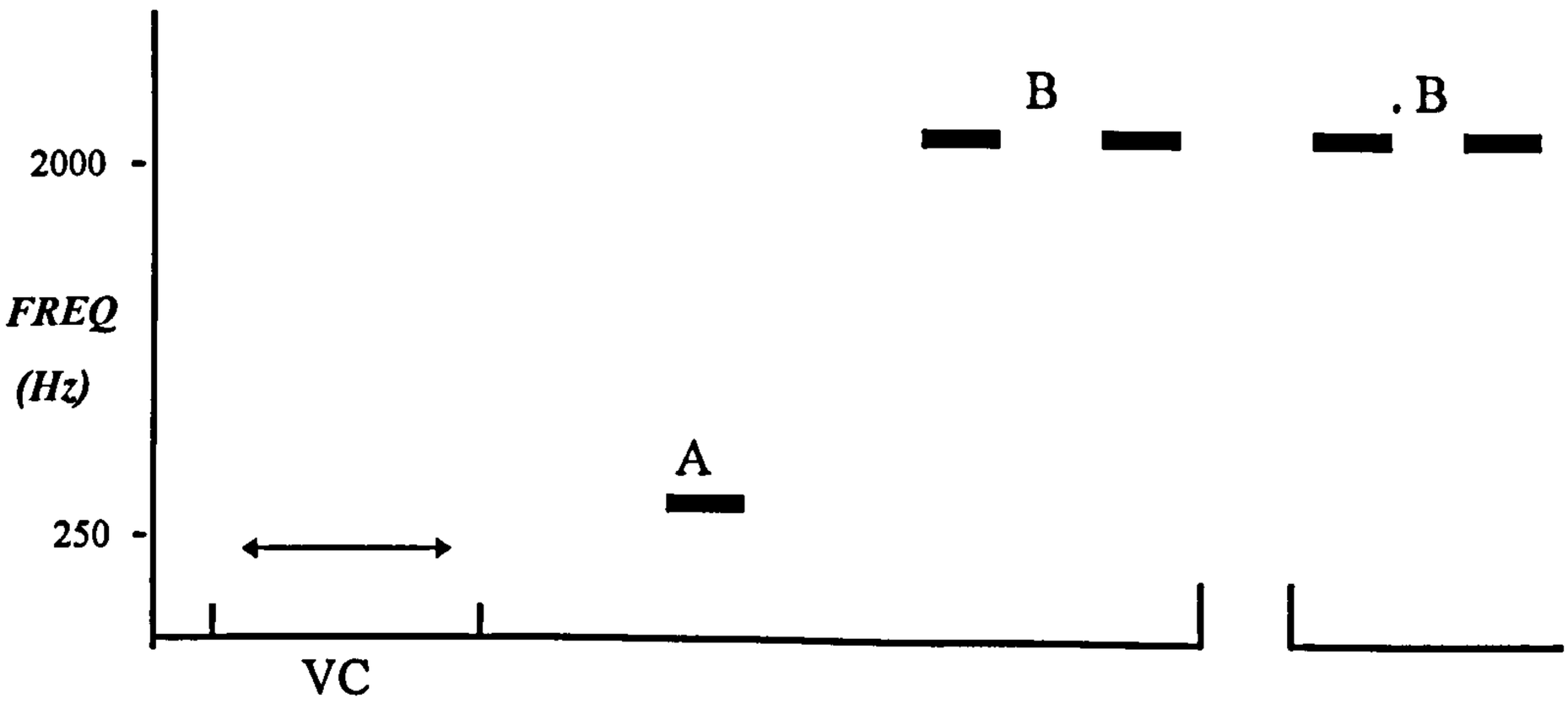
Auditory pointers and stimuli were synthesised using the MITSYN software package (Henke, 1990). Sounds were produced by a CED 1401 laboratory interface controlled by a Dell system 310 PC, and presented to subjects over Sennheiser HD414 headphones. Subjects controlled the position of the pointer with a TRUDOX track-ball with a 4cm ball. The visual pointer and stimuli were presented between the ears of a blue head silhouette on a 640x200 pixel VGA display. Subjects were seated approximately 50cm from the screen in a

darkened sound-attenuating room in all experiments. Head height, orientation and distance from the screen were kept constant with a chin rest. Visual stimuli were presented at eye-level.

3.5 PROCEDURAL OUTLINE

A typical trial, using an auditory stimulus and auditory pointer, is shown in figure 14. Subjects were instructed to wait for the stimulus. If a visual stimulus was to be presented, a 500ms cue to its position preceded the stimulus onset by 500ms. When a visual pointer was used it was presented in a random starting position within the ears of the head silhouette. Visual and auditory pointers were presented 500ms after the offset of the stimulus. Subjects were required to use the track-ball to adjust the pointer until they considered it to be in the position indicated by the stimulus. No time limit was placed on the adjustment procedure.

FIGURE 14



A - auditory stimulus

B - auditory pointer tones

VC - visual cue

The importance of accuracy judgements over speed was highlighted before each session. When subjects had finished their adjustment of the pointer they pressed a key marked "NEXT". The next trial followed after an inter-trial interval of 1500ms.

Variations on these general experimental procedures are detailed in the relevant sections of the chapters that follow.

Chapter 4

4.0 LATERALISATION OF STATIONARY AUDITORY AND VISUAL STIMULI USING AUDITORY AND VISUAL POINTERS.

This experiment had three distinct objectives: (i) To confirm that the relationship between the IID of a 250 Hz tone and its perceived position indicated with a visual pointer was linear over a ±12dBIID range (c.f. Yost 1981). (ii) To investigate to what extent the modality of response modulates the relative influence of the individual modalities when audio-visual stimuli were presented, as suggested by the DAH. (iii) To compare lateralisations made in within-modality conditions (conditions in which both the stimulus and response were either auditory or visual) with lateralisations made in conditions in which stimulus and response modalities differed, thereby establishing a correspondence between visually and auditorily presented positions.

A linear relationship between perceived position and the actual position of the stimulus was expected (c.f. experiment 1, and Yost 1981). The results of experiment 1 suggested that the linear relationship would be independent of whether visual judgements of auditory, or visual stimuli were made, and pilot assessments of different stimuli (general procedure and stimuli, chapter 3)

indicated that judgements of auditory stimuli made with an auditory pointer also show a linear relationship with the IID of the stimulus. The results of the lateral tracking experiment described earlier were consistent with Warren et al's (1983) findings that mean standard deviations could be used as a "tag" in evaluating the relative influence of the modalities in a localisation task. A prediction of the DAH is that mean standard deviations should be characteristic of whether responses had been made auditorily or visually, consistent with the hypothesis that relatively more attention, and therefore relative dominance is allocated to the modality in which the response is made. Similarly it was expected that variance in responses would be influenced by whether stimuli were presented visually or auditorily, again as a function of attention allocation to the modality in which the stimulus was presented (c.f. DAH).

4.1 SUBJECTS

24 subjects took part in the experiment. Pure tone audiometry showed that all subjects had thresholds within the normal range.

4.2 ADDITIONAL PROCEDURAL POINTS

Subjects were instructed to respond to the position of auditory or visual stimuli with auditory or visual pointers in four different conditions. Order of condition presentation was fully counterbalanced. Subjects received a different condition on each day, for four consecutive days. Two practice trials were presented before each session.

4.2.1 Stimuli

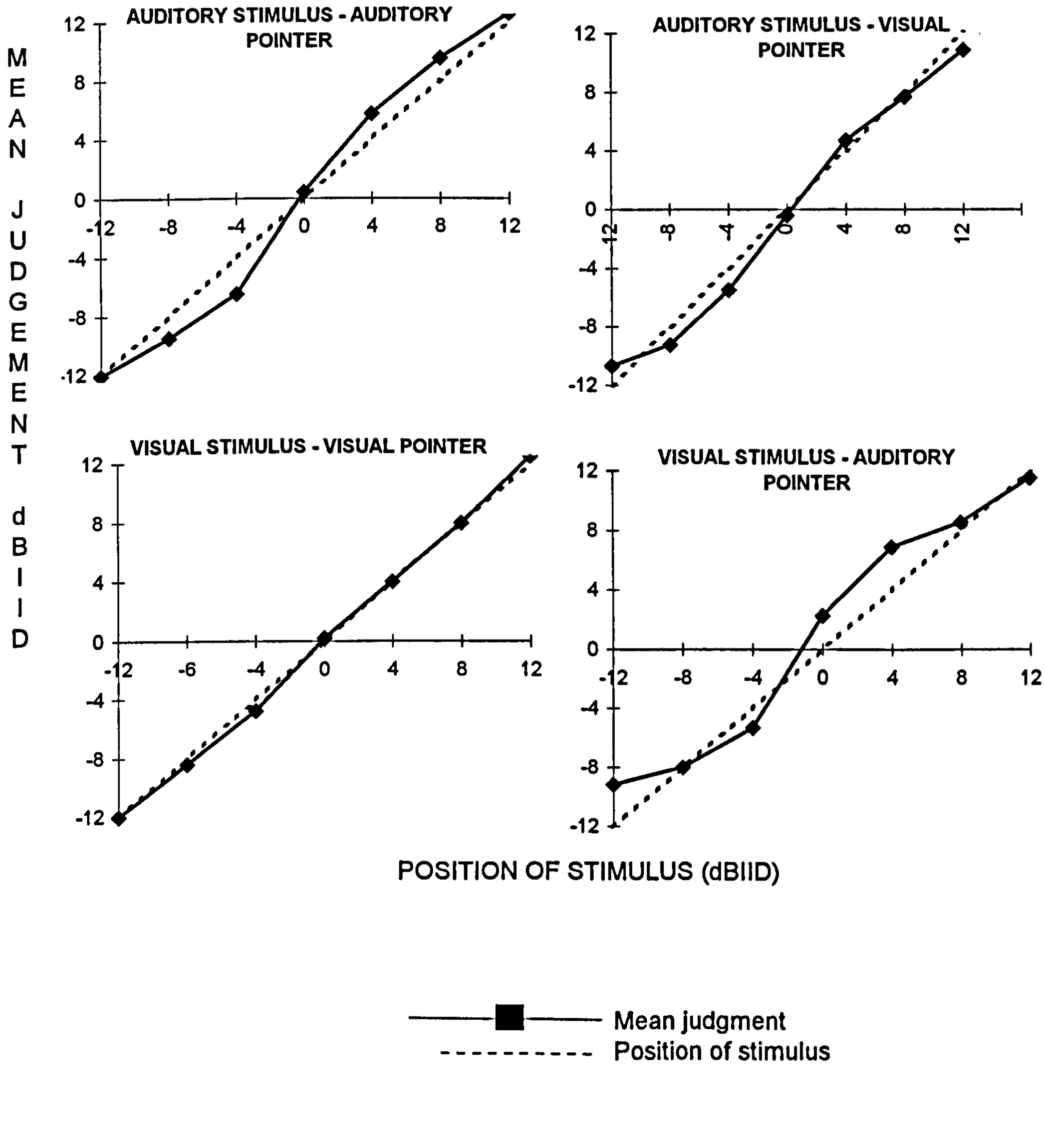
Seven auditory stimuli were drawn from a ±12dBIID range, and differed in gradations of 4dBIID. Seven, equally-spaced visual positions were used, covering a range of 10.6cm. The range of visual stimuli was chosen on the basis of the results of the lateral tracking experiment. 10.6cm (center ± 5.3cm) was the range of the visual pointer used in responses of auditory stimuli in the ±12dBIID range. Otherwise, stimuli were as detailed in chapter 3. Each of the seven positions was tested four times in each condition

4.3 RESULTS

Mean judgements for each of the four conditions can be seen in figure 15. Mean judgements were linearly related to the position of the stimulus. Deviation from the 'correct' line was small in all conditions. Responses with the auditory pointer (lower right-hand panel and upper left-hand panel) tend to be less accurate that those with the visual pointer.

A 3-way analysis of variance with condition (4 levels), stimulus position (7 levels) and repetition of stimulus position (4 levels) as factors showed a significant main effect of stimulus position [$F(6,138)=568.71$, $p<0.001$]. Condition was shown not to be a significant factor [$F(3,69)=1.2$ $p=0.17$]. Repetition of the stimulus was shown not to be a significant factor [$F(3,39)=0.49$, $p=0.69$]. The interaction between condition and stimulus position was not significant [$F(18,414)=1.38$, $p=0.139$].

Figure 15: Mean judgement of lateral position



Standard deviations for each of the subjects were averaged and plotted as a function of stimulus position in figure 16. Mean standard deviations were relatively constant within each condition as a function of stimulus position, but varied in level between conditions. Mean standard deviations in visual judgements of visual stimuli (figure 16 - lower left panel) were consistently lower than mean standard deviations in auditory judgements of auditory stimuli (figure 16 - upper right panel). 2-way analysis of variance with condition (4 levels) and stimulus position (7 levels) as factors showed that stimulus position [$F(6,138)= 1.1$, $p=0.18$] and the interaction between the two factors [$F(18,414)=1.45$, $p=0.08$] were not significant. A significant main effect of condition [$F(3,69)=13.41$, $p<0.001$] was shown. The results of Tukey's HSD aposteriori comparisons are summarised in table 2.

Figure 16: Mean Standard deviations

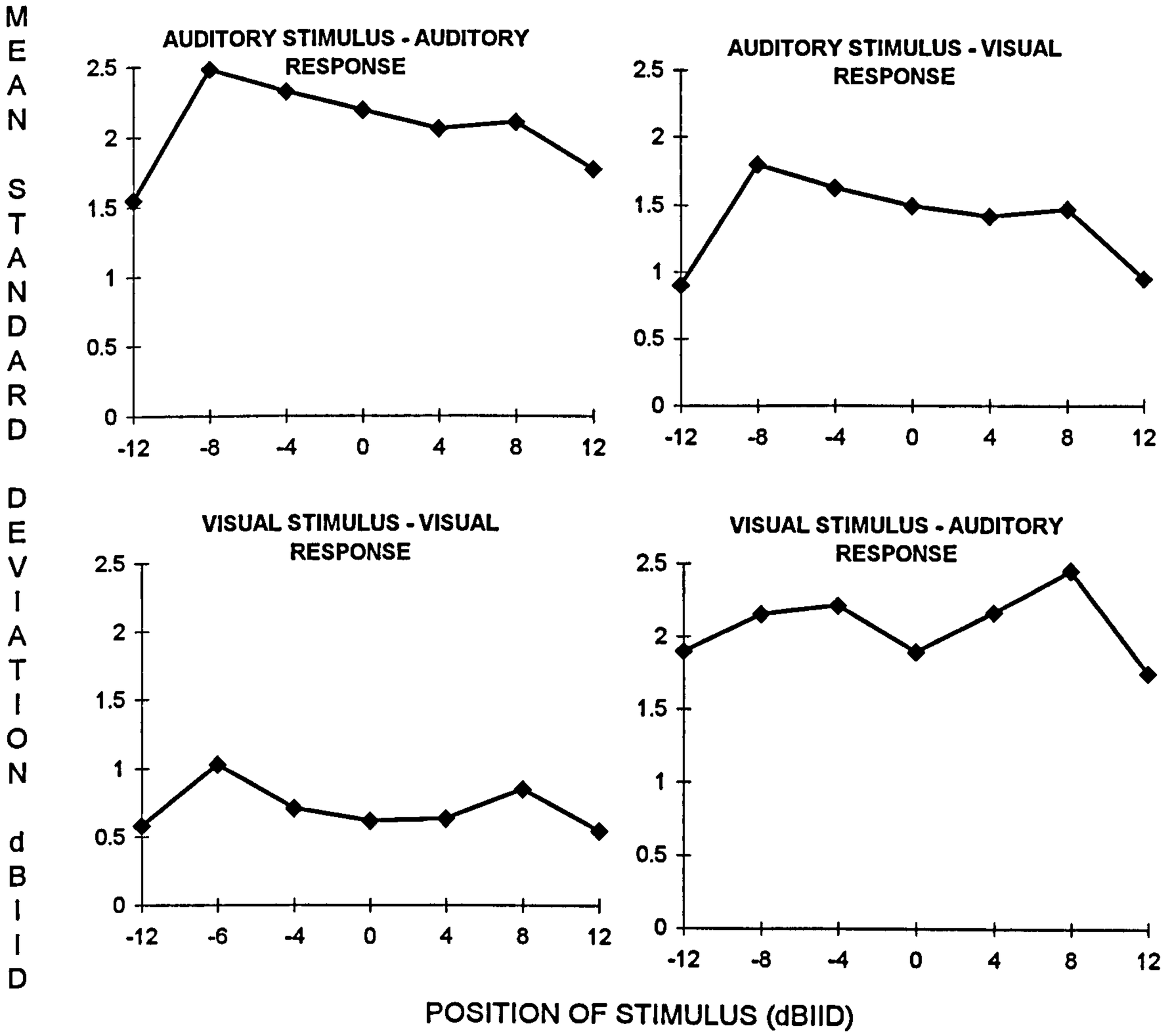


TABLE 2: Tukey's aposteriori comparisons for mean standard deviations in judgements of auditory and visual stimuli using auditory or visual pointers collapsed across stimulus position. HSD p<0.01=0.94; HSD p<0.05=0.766

	AUD. STIMULUS AUD. RESPONSE mean = 2.43	**VIS. STIMULUS VIS. RESPONSE** mean = 0.73	**AUD. STIMULUS VIS. RESPONSE** mean = 1.51
VIS. STIMULUS VIS. RESPONSE mean = 0.73	Difference = 1.7 Sig. p<0.01	XX	Difference = 0.75 Non-Sig.
AUD. STIMULUS VIS. RESPONSE mean = 1.48	Difference =0.95 Sig. P<0.01	Difference = 0.75 Non-Sig.	XX
VIS. STIMULUS AUD. RESPONSE mean = 2.07	Difference =0.36 Non-Sig.	Difference = 1.34 Sig. p<0.01	Difference = 0.56 Non-Sig

4.4 DISCUSSION

Mean judgements of lateral position (figure 15) in all conditions were consistent with the hypothesis that there would be a linear relationship between perceived position and stimulus position (c.f. exploratory studies made in the general procedure section, and those of Yost 1981). The data showed that the perceived position of a stimulus presented in one modality was independent of whether subjects indicated the perceived position with an auditory or visual pointer. Similarly, the perceived position of a stimulus indicated visually or auditorily was independent of whether or not the stimulus

was presented in the same modality as the pointer. The data also established a correspondence between visually and auditorily presented lateral positions, which was a prerequisite for the presentation in later experiments of audio-visual stimuli with laterally-corresponding modal components.

A ceiling effect has been discussed as a possible factor in the results of the preliminary tracking experiment described in chapter 2. It is possible that the data shown in figure 15 could be explained similarly. It is possible that the task was too easy, and that making the task more difficult might reveal an effect of condition in the mean position data. However, an explanation of the apparent non-effect of condition on mean lateralisation judgements purely in terms of a ceiling effect seems unlikely. Trial to trial improvement on the task was investigated in terms of comparison of judgements made on each repetition of each stimulus position. No trial to trial improvement was shown which suggests that if subjects had reached ceiling they did so immediately. Two practice trials were presented before each session, but it seems unlikely that this minimal prior experience with the stimuli provided subjects with anything other than a glimpse at the task.

Response accuracy, measured in terms of mean standard deviations did not differ as a function of stimulus position, but did differ as a function of condition. Aposteriori tests (table 1) showed that the differences were partly a function of response modality.

- For a given response modality, response accuracy did not differ between auditory and visual stimuli

- Mean standard deviations were significantly larger in the auditory within modality condition than mean standard deviations in the visual within modality condition.

- Comparison of conditions within which both the stimulus and response modalities differed (right-hand panels of figure 16) were not significant, suggesting that judgement accuracy was not simply a function of response modality.

When stimulus and response modalities differed, mean standard deviations were influenced by the modality of response. That is to say, when the response was auditory, mean standard deviations were not significantly different to those in the auditory within modality condition (auditory stimulus and response), and when the response was visual, mean standard deviations were not significantly different to those in the within visual condition (visual stimulus and response) This is consistent with the predictions of the DAH. The DAH suggests that the relatively dominant modality will be that to which relatively more attention has been directed. It was predicted earlier that the direction of attention in this sense might be afforded by the identity of the response modality, i.e. a visual response would direct attention to the visual modality and an auditory response might direct attention to the auditory

modality. Over all, the results suggested that judgement accuracy was a function of response modality and also a function of stimulus modality.

4.5 SUMMARY and CONCLUSIONS

Mean judgements of the position of stimuli were independent of stimulus and response modalities (c.f. preliminary lateral tracking task, chapter 2). The data demonstrate a correspondence between visually and auditorily presented lateral positions, allowing the presentation in later experiments of audio-visual stimuli with laterally corresponding, or non-corresponding components, in which the non-correspondence could be systematically manipulated.

Response accuracy, measured as mean standard deviations, differed as a function of response modality, although stimulus modality may also affect accuracy when stimulus and response are in different modalities. Mean standard deviation in responses has been shown to be a useful tool in discriminating between judgements on tasks based entirely in the visual or auditory modality (c.f. Warren et al 1983).

4.6 IMPLICATIONS

The results suggested than response accuracy (measured in terms of mean standard deviations) is likely to be more informative than response bias (measured in terms of mean judgements of lateral position) regarding whether responses are based auditorily or visually. It appears that mean standard deviations can be used as a "tag" (c.f. Warren et al 1983) in investigations of

whether judgements of audio-visual stimuli are influenced more by information provided in the visual modality, or by information provided in the auditory modality. The influence of response modality on response variance motivated the choice of a single, consistent response modality – the auditory pointer – in subsequent experiments.

Chapter 5

5.0 LATERALISATION OF AUDIO-VISUAL STIMULI WITH SPATIALLY-CORRESPONDING MODAL COMPONENTS.

Mean lateralisation judgements of stimuli used in the previous experiment were independent of both the stimulus modality and the modality in which responses were made. The relationship between stimulus position and perceived position indicated with a visual pointer or with an auditory pointer was approximately linear up to lateral positions of ±12 dBIID (c.f. Yost 1981). Comparison of auditory responses to visual stimuli and auditory responses to auditory stimuli established a correspondence between visually and auditorily-presented lateral positions. This makes possible the presentation of audio-visual stimuli with auditory and visual components which can be said to correspond in perceived lateral position.

It was the objective of this experiment to investigate whether judgements of audio-visual stimuli differed from judgements of either visual or auditory stimuli. Mean standard deviation in judgements was chosen as the metric for comparing the relative accuracy of judgements of auditory, visual and audio-visual stimuli. Previous experiments have shown that the comprehension of a speech signal can be improved if the listener can see the face of the talker (Sumby and Pollack 1954; Macleod and Summerfield 1990; Plomp and

Mimpen 1979). In a noisy environment, increasing the amount of perceptual information available to the listener by providing corresponding visual information (moving lips) can enhance the perception of speech (Cherry 1953). An accompanying visual stimulus can aid in disambiguating a complex auditory environment. In these cases, judgements based on corresponding information provided in two modalities were more accurate than judgements of information provided in either the auditory or visual modality alone.

This experiment involved a comparison of lateralisation judgements of bi-modal audio-visual stimuli with judgements of uni-modal, auditory or visual stimuli. The experiment described in chapter 4 showed that mean accuracy of lateralisation judgements of the stimuli used was partly a function of the response modality. In this experiment subjects responded with an auditory pointer in all conditions, so as to control for any possible influence of response modality.

Subjects were given carefully-worded instructions regarding their task in the experiment. It has been suggested that experimental instructions can have a direct influence on the subjects' assumption of the 'unitariness' of the perceptual event (Welch and Warren 1980). The Directed Attention Hypothesis (DAH) suggests that relative dominance is exerted by the modality towards which relatively more attention is directed. Instructing subjects to rely on one modality relative to another has been used to manipulate this allocation of attention (Pick et al. 1969; Warren and Schmitt 1978). Warren et

al. (1981) showed that the distance over which subjects would relocate a voice to a spatially separated moving mouth (the ventriloquist effect) could be influenced by specifically instructing subjects that the auditory and visual stimuli they were to receive referred to the same event (unitary-event instruction). The wording of the instructions here was chosen in an attempt to avoid biasing one modality in favour of the other (appendix I).

It was hypothesised that the relationship between mean judgements of lateral position and stimulus position would be approximately linear independent of whether stimuli were presented in the auditory modality (condition 1) or visual modality (condition 2) c.f. Chapter 4. There was no reason to suspect that a similar relationship would not extend to auditory judgements of audio-visual stimuli.

It was hypothesised that mean standard deviations in conditions 1 and 2 would not differ significantly (c.f. chapter 4). The results of the experiment described in chapter 4 indicated that the accuracy of lateralisation judgements (measured in mean standard deviations) was primarily a function of response modality. Since auditory judgements of lateral position were made in all three conditions, means and standard deviations could not differ as a function of response modality. It follows that any differences in mean standard deviations would be a function of stimulus type.

5.1 STIMULI

The stimuli used in the previous experiment were used here. Audio-visual stimuli were spatially-corresponding auditory and visual stimuli. The auditory and visual components of the audio-visual stimuli were presented simultaneously, making them temporally and spatially correspondent.

5.2 RESPONSE

Subjects responded with the auditory pointer described earlier.

5.3 SUBJECTS

Twelve subjects took part in the experiment. All subjects had previously taken part in the experiment described in chapter 4.

5.4 PROCEDURE

Three conditions were presented to each subject. In condition 1, auditory stimuli were presented, in condition 2 visual stimuli, and in condition 3 subjects responded to audio-visual stimuli. Subjects were presented with a different condition on each day for three consecutive days. The order of condition presentation was fully counterbalanced. Each of the seven positions were tested four times in each condition. Subjects were asked to read through the instructions before each session (appendix I).

5.5 RESULTS

Mean judgements as a function of stimulus position are plotted in figure 17. A linear relationship between perceived position and stimulus position is shown, with some deviation from linearity noticeable particularly in condition 3 at more extreme lateral positions. A 3-way analysis of variance with condition (3 levels), stimulus position (7 levels) and stimulus repetition number (4 levels) was performed. The main effects of condition [$F(2,20)=0.18$, $p=0.84$)] and stimulus repetition number [$F(3,30)=1.71$, $p=0.186$] were not significant. Stimulus position was a significant factor in the analysis [$F(6,60)=125.33$, $p<0.001$]. The interaction between condition and stimulus position was also shown to be significant [$F(12,120)=1.92$, $p=0.039$).

Mean standard deviations in each condition are plotted in figure 18. A 2-way analysis of variance with condition (3 levels) and stimulus position (7 levels) as factors showed a significant main effect of condition [$F(2,20)=5.16$, $p=0.016$] but stimulus position [$F(6,60)=1.63$), $p=0.154$] was not a significant factor. The interaction between condition and stimulus position was not significant [$F(12,120)=1.04$, $p=0.42$]. Tukey's aposteriori comparisons (summarised in table 3) indicated that standard deviations in condition 3 were significantly smaller than standard deviations in condition 1, but not significantly different to standard deviations in condition 2. Standard deviations in conditions 1 and 2 were not found to be significantly different.

FIGURE 17: MEAN JUDGEMENTS OF LATERAL POSITION

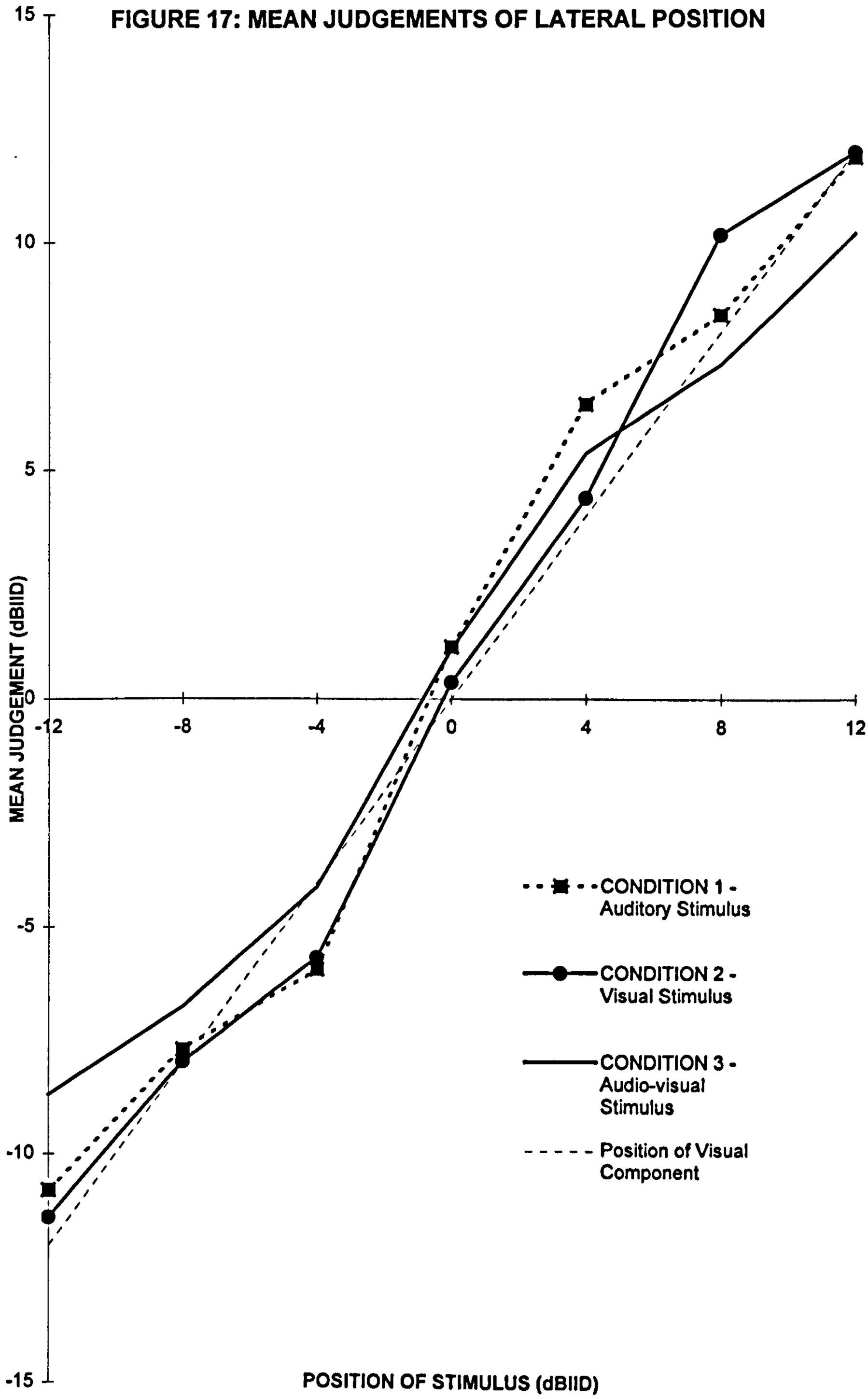


FIGURE 18: MEAN STANDARD DEVIATIONS IN LATERALISATION JUDGEMENTS

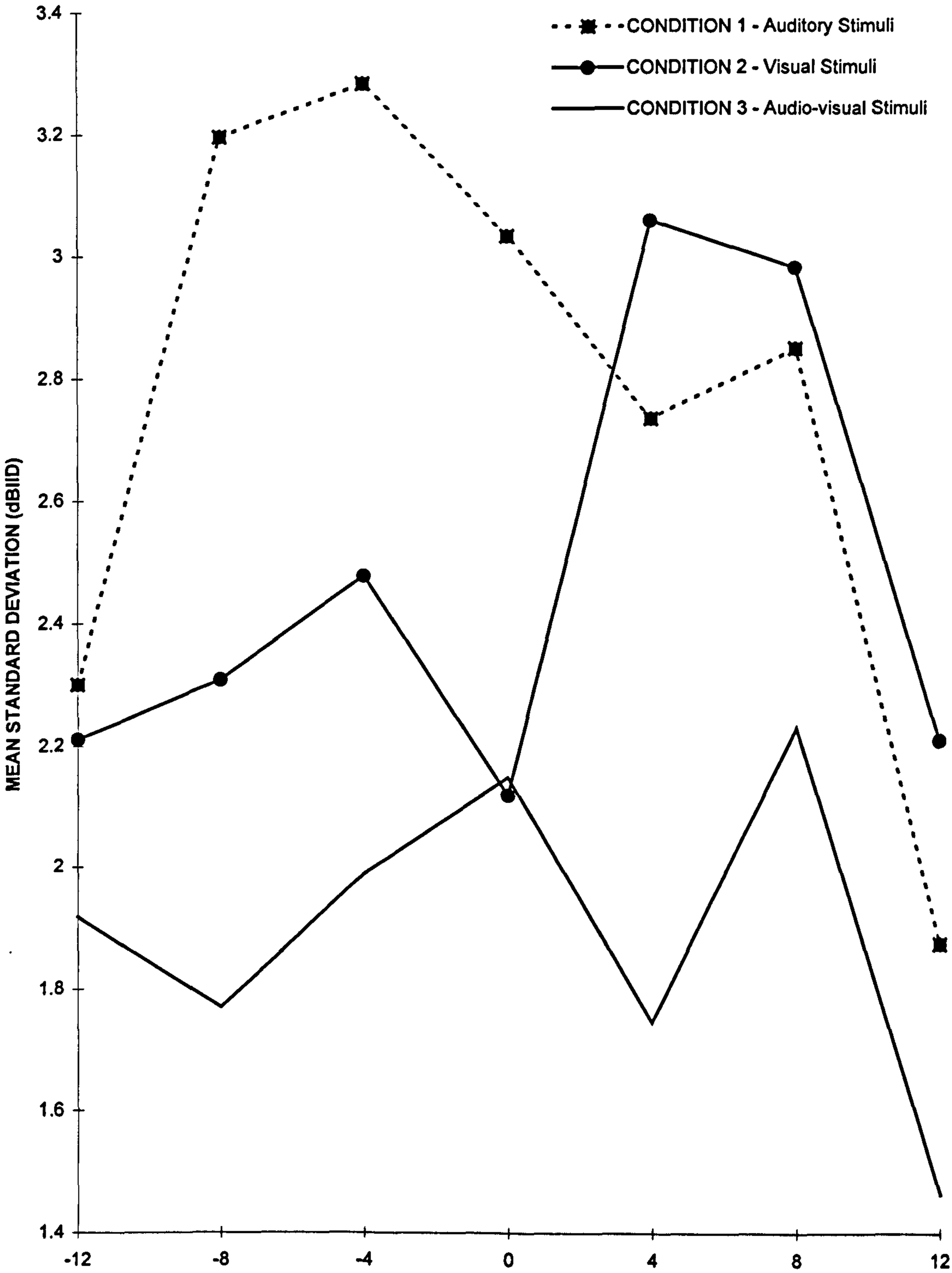


TABLE 3 Tukey's HSD comparisons of mean standard deviations collapsed over stimulus position as a function of condition in experiment 2. Condition 1=Auditory stimuli; Condition 2=Visual stimuli; Condition 3=Audio-visual stimuli. HSD $p<0.01$=0.856dBIID, HSD $p<0.05$=0.6532.dBIID

	CONDITION 1 mean = 2.752	CONDITION 2 mean = 2.48
CONDITION 2 mean = 2.48	Difference = 0.272dBIID Not Significant	xx
CONDITION 3 mean = 1.89	Difference = 0.862 dBIID Significant: $p<0.01$	Difference =0.59 dBIID Not Significant

5.6 DISCUSSION

Mean judgements of lateral position - shown in figure 17 - were consistent with hypotheses and predictions made earlier. Mean judgements of auditory and visual stimuli did not differ significantly. Mean judgements of audio-visual stimuli did not differ significantly from mean judgements of auditory or visual stimuli, as expected. Mean judgements in all conditions were approximately linearly related to the position of the stimulus (c.f. chapter 4 and Yost 1981). Analysis of variance showed a significant interaction between presentation position and condition. The significant interaction was likely to be a result of differences in linearity as a function of condition (figure 17). The data suggests that mean judgements of audio-visual stimuli were compressed into a narrower range than judgements of auditory or visual stimuli alone. Data from the previous experiment (figure 15) indicated that the

relationship between mean auditory lateralisation judgements and the lateral position of auditory or visual stimuli showed some non-linearity. Because of this, some deviation from linearity in the relationship between auditory judgements of audio-visual stimuli and the lateral position of the stimuli was not unexpected. It is possible that the deviation from the 'correct' line in these data, and the loss of linearity in mean judgements of extreme auditory and visual positions was compounded when audio-visual stimuli were presented. However, and most importantly in the context of this experiment, no differences in mean lateralisation judgements as a function of stimulus type were found.

Mean standard deviations in judgements were used as a metric for measuring mean judgement accuracy. Consistent with the data reported in chapter 4, mean standard deviations in judgements of auditory and visual stimuli did not differ significantly. Mean standard deviations in judgements of audio-visual stimuli were significantly smaller than mean standard deviations in judgements of auditory stimuli, but not significantly smaller than those of judgements of visual stimuli. Using mean standard deviation as a "tag" (c.f. chapter 4, and Warren et al 1983) the results can be interpreted as indicating a relative dominance of the visual modality in judgements of the audio-visual stimuli presented here. However, although the difference between conditions 2 and 3 was not statistically significant, as figure 18 shows, the variance in judgements of audio-visual stimuli was numerically smaller than the variance in judgments of auditory or visual stimuli at six of the seven positions tested

which suggests that mean standard deviations in judgements of audio-visual stimuli were consistently smaller than mean standard deviations in judgements of auditory or visual stimuli. Mean judgements of lateral position (figure 17) indicated no significant difference in mean judgements of auditory, visual and audio-visual stimuli. Combined with the mean standard deviation data, this suggests that mean lateralisation judgements of audio-visual stimuli were consistently more accurate than mean lateralisation judgements of auditory or visual stimuli.

Figure 17 also suggests a 'center' effect for mean judgements of visual stimuli, with mean judgements close to the correct position when central stimuli were presented. This is not as clear for mean judgements of auditory or audio-visual stimuli. Figure 18 shows a corresponding dip in mean standard deviations for judgements of visual stimuli at the central position which was not found for judgements of auditory or audio-visual stimuli. This observation corresponds with the accuracy with which subjects can bisect lines, although it is not clear why a corresponding 'bisection effect' was not seen for audio-visual, and possibly auditory stimuli. Visual stimuli were presented on an axis between the ears of a head silhouette. If a stimulus was presented in a central position - equivalent to 0dBIID - bisecting the axis would give an accurate lateralisation judgement. Roig and Cicero (1994) showed that the average error in bisecting lines ranging between 26mm and 111mm in length was 0.44mm. The visual stimuli in this experiment were approximately 2mm in

width, and a line bisection accuracy of the magnitude suggested by Roig and Cicero would allow subjects to lateralise the stimulus with great precision.

5.7 SUMMARY and CONCLUSIONS

The results indicated that there were no significant differences between mean lateralisation judgements of auditory, visual and audio-visual stimuli, (c.f. preliminary data described in chapter 2), and that there was an approximately linear relationship between stimulus position and perceived position independent of stimulus modality. Mean standard deviations in judgements of the lateral position of audio-visual stimuli were significantly smaller than mean standard deviations in judgements of auditory stimuli. The results suggest that stimulus modality is a factor in the mean accuracy of lateralisation judgements, but not mean lateralisation judgements. Using mean standard deviation as a "tag" (c.f. Warren et al 1983) the results show a relative dominance of the visual stimulus in the accuracy of lateralisation judgements of audio-visual stimuli. However, the data also suggest an influence of both the auditory and visual modalities in lateralisations of audio-visual stimuli, with mean accuracy in judgements being numerically greatest for judgements of audio-visual stimuli than judgements of auditory or visual stimuli.

5.8 IMPLICATIONS

A slight flattening of mean judgement curves (figure 17) at IIDs greater than ± 8dBIID indicates some degree of non-linearity. For this reason, stimuli within

the ±8dBIID range should be used to ensure the linear relationship between stimulus and perceived positions.

Chapter 6

6.0 THE AUDIO-VISUAL SPATIAL RELATIONSHIP.

The results of the previous experiment suggested that lateralisation judgements of audio-visual stimuli were influenced by both the auditory and visual modalities. Mean standard deviations of judgements of audio-visual stimuli were numerically smaller than the mean standard deviations of judgements of auditory or visual stimuli. This implies that subjects' lateralisation estimates of audio-visual stimuli were based on a combination of auditory and visual spatial information rather than visual or auditory information alone.

The results reported so far suggest a relative dominance of information in the visual modality, although the results of the experiment described in chapter 5 showed that the mean accuracy in judgements of audio-visual stimuli was greater than the mean accuracy in judgements of auditory or visual stimuli. This experiment was an investigation of the effects on perceived lateral position of spatially mis-matching the auditory and visual components of an audio-visual stimulus. The results of the experiment were intended to provide insights into the relative importance placed on auditory and visual information in this lateralisation task.

The relative dominance of the visual modality in tasks requiring the localisation of audio-visual stimuli is well documented. Pseudophones have been used to alter the apparent source of the auditory component of an audio-visual stimulus (Young 1928; Willey et al. 1937 cited by Welch and Warren 1980; Held 1955). A series of pipes and horns transferred the sound which would normally enter one ear to the other. Young (1928) found that in the absence of the visual component of an audio-visual stimulus, subjects perceived a change in the apparent position of the auditory image due to the action of the pseudophone. When the visual component was reintroduced, the apparent position of the auditory component relocated to the position of the visual component. In this case, when the spatial information provided by the auditory and visual components did not correspond, the perception was of a common source in the position of the visual component. This perceptual relocation of sound to the position of a visual stimulus, or ventriloquism, has also been shown with non-speech stimuli. Subject showed a propensity to perceive the apparent source of a steam whistle as being in the position of a simultaneously presented jet of steam although the two components were in fact spatially separated (Jackson 1953). Jackson went on to measure the distance the auditory and visual components could be separated before the sound was no longer relocated to the visual component - the level of ventriloquism, was partly dependent on the context of the audio-visual pairing. He showed that the level of ventriloquism for "less-meaningful" audio-visual pairings was less than for "more-meaningful" pairings. The sound of a bell and a light showed less ventriloquism than the steam/steam-

whistle pairing because, it was claimed, subjects were more familiar with the steam/steam-whistle pairing than the light/bell pairing. Similar results were shown by Thurlow and Jack (1973) and Radeau and Bertelson (1977). Both showed that the contextual realism of an audio-visual event affected the level of ventriloquism associated with it.

In all cases, a dominance of the visual modality was shown. When the auditory and visual components of an audio-visual stimulus were made spatially discrepant, within a certain range, mean judgements of spatial position were always in the position of the visual component. Relative visual dominance, in the experiments cited here, is predicted by the Modality Appropriateness Hypothesis (MAH), the Modality Precision Hypothesis (MPH), and the Directed Attention Hypothesis (DAH). However, it is not clear whether subjects' judgements of localisation were based entirely on the visual component, and what influence, if any, the auditory components of the stimuli had on subjects' perceptions of their apparent source.

Jackson (1953) suggested that the identity of the auditory component affected the level of ventriloquism. If the auditory component had influenced subjects' judgements it could have been as a general distracter, in which case the position of the auditory component relative to the position of the visual component would be irrelevant. If it were simply the presence of an auditory stimulus in the task that was distracting the subject, the level of distraction, or influence of the auditory component in the task should be constant for all

levels of audio-visual spatial discrepancy. However, if subjects make use of information about the position of the auditory component, its influence on their judgements of audio-visual location should vary with the level of audio-visual spatial discrepancy. This is suggested by Jackson (1953) who also showed that the proportion of responses indicating that the sound seemed to emanate from the auditory source rather than the visual source increased as a function of audio-visual spatial mismatch, although the level of ventriloquism was high even at relatively large audio-visual separations. In summary, a judgement based on the visual component, because of its dominance, appropriateness, and relative precision in spatial tasks (c.f. MPH, MAH, DAH), may have been influenced by the relationship between the auditory and visual components and the spatial separation between the components. The very presence of the auditory component may also have acted as a distracter in what might otherwise have been a simple visual localisation task.

The objective of this experiment was to investigate the effects of spatially mismatching the auditory and visual components of audio-visual stimuli on subjects' judgements of lateral position. Thresholds for audio-visual spatial mismatch were determined initially to allow an assessment of the effect on lateralisation judgements of the detectability of audio-visual spatial non-correspondence.

6.1 MEASUREMENT OF AN AUDIO-VISUAL SPATIAL-CORRESPONDENCE DIFFERENCE LIMEN.

6.1.1 STIMULI

Audio-visual stimuli were presented with synchronous auditory and visual components which differed spatially by varying amounts.

6.1.1.a Visual Components

Visual stimuli (detailed in Chapter 3) were presented in two possible positions on a 22cm axis drawn between the ears of a head silhouette. The visual positions were analogous to auditory positions of ±3dBIID.

6.1.1.b Auditory Components

Tones (detailed in Chapter 3) could be presented in lateral positions in the range ±7dBIID in 1dB steps relative to the position of a visual component. Tones were centered on the positions of the visual component, equivalent to interaural intensity differences of -3dBIID (lateralised on the left of intracranial center - ICC), or +3dBIID (lateralised on the right of ICC) depending on the position of the visual component being tested (see diagram).

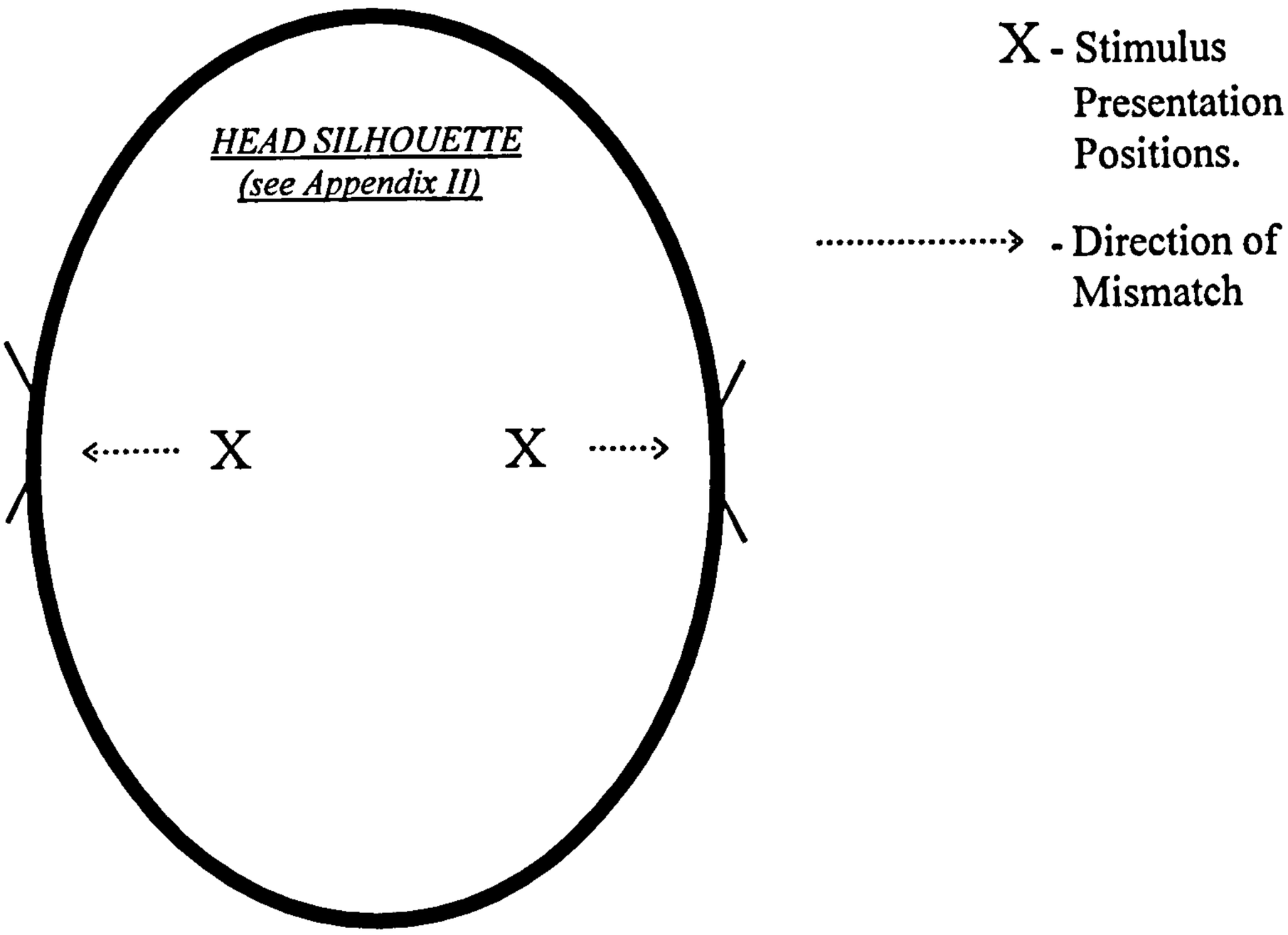
6.1.2 SUBJECTS

Fifteen subjects took part in the experiment. Four had taken part in previous experiments. Audiometric tests showed that all the subjects had pure-tone thresholds within the normal range.

6.1.3 PROCEDURE

Pairs of audio-visual stimuli separated by 100ms were presented in a 2I-2AFC procedure. A 100ms visual cue in the position that the visual component of the stimulus was to be presented preceded each trial by 100ms (see diagram). Subjects were required to indicate the interval in which the auditory and visual components were spatially correspondent (the target) by pushing one of a pair of keys marked 1 and 2. The interval in which the target was presented was determined randomly, with an equal number of target presentations in the first and second intervals. The next trial was presented after a two second inter-trial delay.

A practice session of 30 trials was presented before each experimental session. The 30 possible configurations of audio-visual spatial mismatch (visual position ±7dBIID) were presented once each in random order. Subjects were provided with feedback during the practice session but not during the experimental sessions.



Stimuli presented in positions indicated by 'X' representing + and - 3dBIID.

The 30 spatial mismatch possibilities was presented forty times each in random order. Presentation order was randomised with the constraint that each mismatch possibility would be presented an equal number of times relative to each visual position. Breaks were given after every 150 trials.

6.1.4 RESULTS

Mean errors for all subjects as a function of auditory/visual spatial mismatch are plotted in figures 19 and 20. Figure 19 shows mean errors for stimuli with visual components in position -3dBIID, that is on the left hand side of the head. Figure 20 shows mean errors for stimuli with visual components on the right hand side of the head, in position +3dBIID. Maximum errors occurred with stimuli differing spatially by +1dBIID for left presentations, and -1dBIID

for right presentations. In both cases, maximum errors were obtained when the auditory component of the audio-visual stimulus was spatially mismatched from the visual component towards ICC (intracranial center). Smaller errors were found with spatial mismatches of a corresponding size in the opposite direction, away from ICC. Both functions show a fairly smooth decrease in error rates as a function of audio-visual spatial mismatch, although error rates are still fairly high (approximately 15%) even at the greatest levels of mismatch.

FIGURE 19: Mean Errors in Judgements of Stimuli with Auditory Components Mismatched Relative to a Visual Component at -3dBIID. (ICC at +3cBIID mismatch)

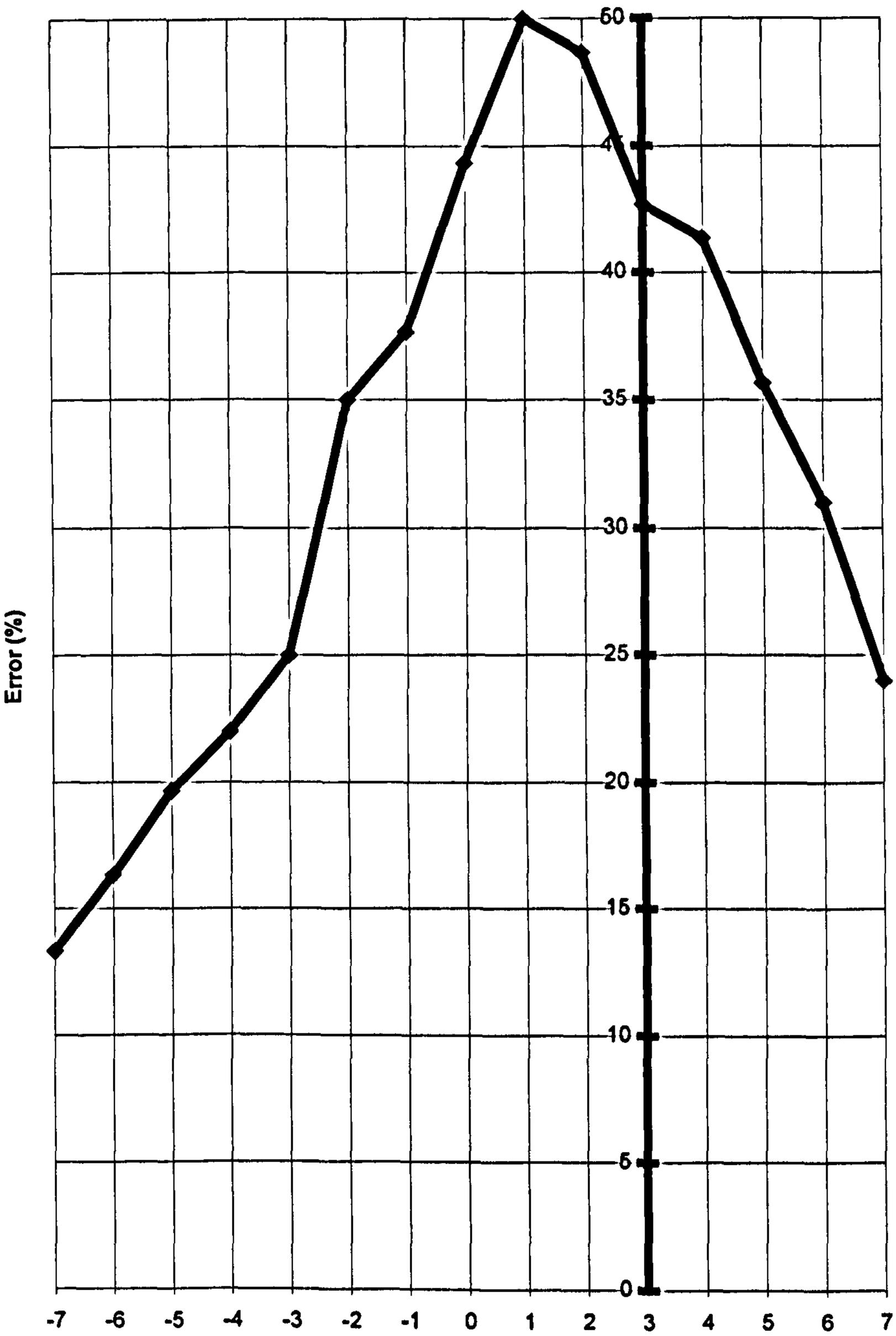
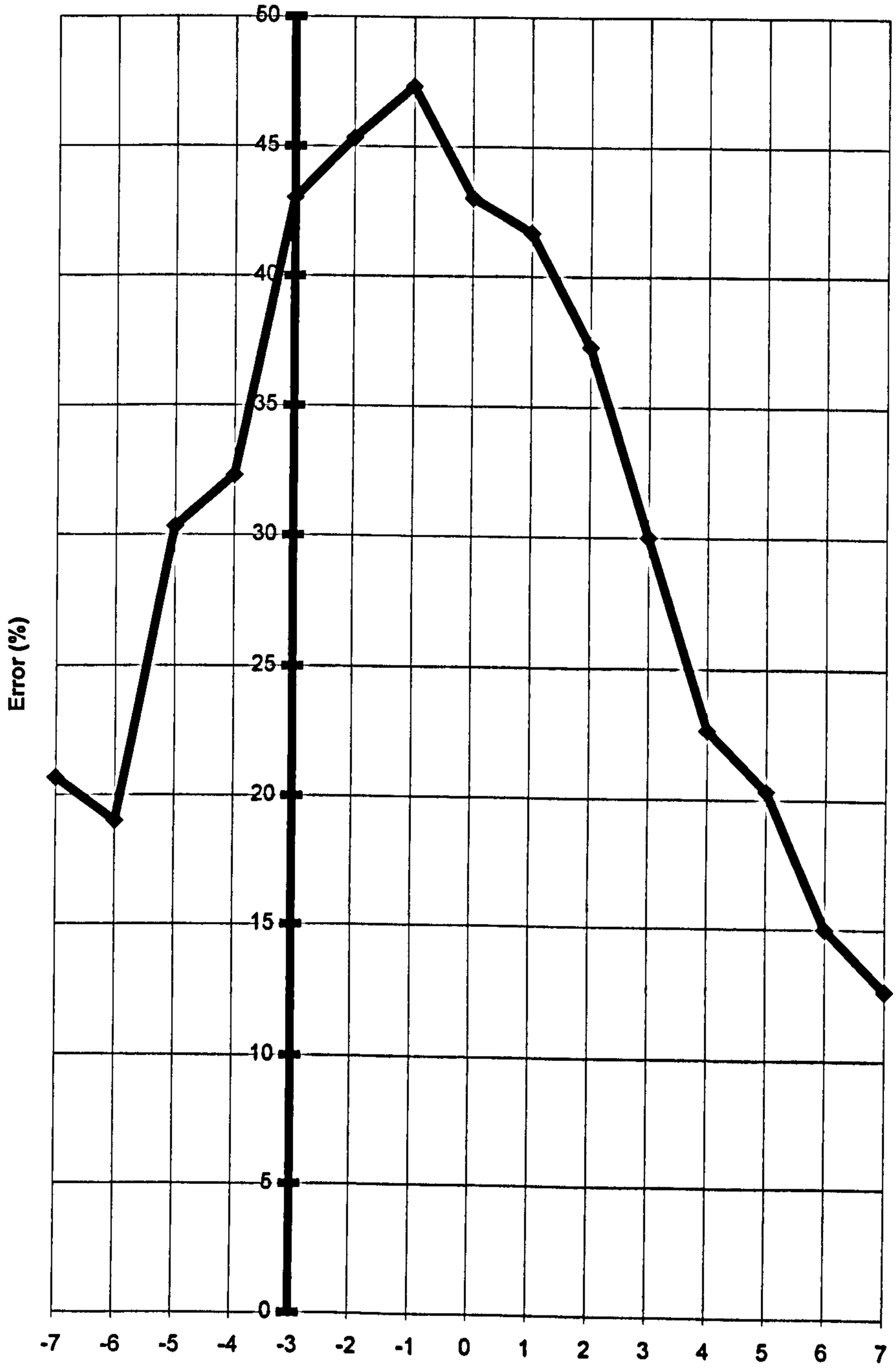


FIGURE 20: Mean Errors in Judgements of Stimuli with Auditory Components Mismatched Relative to a Visual Component at +3dBIID. (ICC at -3cBIID mismatch)



6.1.5 DISCUSSION

The data showed an inverse relationship between audio-visual mismatch and error rate, with a maximum in error rates when auditory components were mismatched away from the position of the visual component towards the ICC.

A 2I-2AFC procedure was employed whereby subjects were required to choose between two audio-visual stimuli, indicating the stimulus with spatially corresponding auditory and visual components. In each trial, one of the stimuli presented had spatially consistent modal components, the other had spatially mis-matched components with the exception of the case when a mismatch level of 0dBIID was presented. If these two physically identical stimuli with spatially matching components were perceptually identical an error rate of approximately 50% would be expected. Figures 19 and 20 show that at 0dbIID spatial mismatch (ICC) subjects produced average error rates of 43.7%. Introducing a spatial mismatch in the modal components of one of the alternatives in the 2AFC should have made the task easier, with maximum error rates expected when the stimuli in the two intervals were minimally discriminable. When the auditory component was displaced outwards, towards the leading ear, the effect on error rate was as expected, performance improved. However, displacing the auditory component inwards, to a position between the visual component and ICC, caused an increase in error rate. The results indicate that it is not simply a difficulty in detecting the spatial mismatch towards ICC. It is not clear why an audio-visual stimulus with an

auditory component mismatched towards ICC by approximately 1dBIID relative to the position of the visual component should be chosen as more likely to indicate audio-visual spatial consistency than an audio-visual stimulus with spatially correspondent components. It may be that an stimulus with an auditory component at, or near ICC is likely to indicate audio-visual spatial correspondence based on past experience. The role of the auditory system in the localisation of sounding, visual objects is partly attention-directing. If an audio-visual stimulus is not within the visual field, the auditory system provides spatial information about the object that enables the listener to look towards the stimulus, aiding its accurate localisation and identification. Both auditory and visual information regarding the stimulus are then available, which may allow the subject to make more accurate judgements of the stimulus than if only auditory or visual information had been available (c.f. Cherry 1953). This head-turning reflex, or orienting reflex is shown in very young infants, who show orientation to sounds in light and dark conditions (Morrongiello 1994). Bower (1982) says that the orienting reflex provides the infant with guaranteed examples of audio-visual spatial correspondence at a stage when the auditory and visual maps are still developing. These experiences probably serve to align the auditory and visual maps in the superior colliculus (c.f. Stein et al. 1994). As the infant develops, the importance put on audio-visual spatial correspondence is reduced. Experiments with ventriloquism have shown that the difference between spatial positions of sound and vision often has to be quite considerable before adults even notice it (Radeau and Bertelson 1982).

Adults do, however, show the orienting reflex. The result of turning the head and facing the sounding object is to bring the level in the two ears into equilibrium. These experiences have probably shown listeners that audio-visual spatial correspondence is partly indicated by an auditory component at the ICC, with a corresponding binaural level balance. It is possible that when faced with a difficult choice, subjects' experience suggests that the alternative with the auditory component nearer ICC is most likely to have spatially correspondent auditory and visual components. A similar strategy may have been employed in trials where the incorrect alternative had an auditory component mismatched relative to the visual component away from the ICC. If the mismatch was very small and the choice was not immediately clear, subjects again chose the alternative with the auditory component nearest the ICC, in this case the correct choice.

The lateral position of the visual components, and therefore the eccentricity of subjects' gaze, may have influenced the perceived lateral position of the auditory component of the stimulus. Gopher (1973) showed a tendency for subjects to look in the direction in which they were listening, and goes on to suggest that eye position may be a guide to the allocation of attention in a particular direction. Reisberg et al (1981) showed that selective listening was influenced by eye-position. Most relevant in the context of this experiment, Lewald and Ehrnstein (1996) showed that the interaural intensity difference - IID- at which a 2kHz tone was perceived as being on the auditory medial plane

was a function of gaze direction. If subjects gaze was directed 45° to the right, the IID at which the tone appeared to be on the auditory medial plane was shifted to the left. Similarly, when gaze was directed 45° to the left, the IID at which the tone appeared to be on the medial plane was shifed to the right. The result suggests that when gaze was eccentrically directed, a tone with an IID normally indicating a central position (0dBIID) was shifted in the direction of the gaze. In order to place the tone back on the medial plane, subjects compensated for the effect of gaze direction by adjusting the IID of the auditory stimulus in favour of the ear opposite to the eccentricity of their gaze. An explanation for the position of maximum errors in the detection of audio-visual spatial mismatch shown in figures 19 and 20 can be offered in these terms. The data shown in figure 19 refer to judgements of stimuli with the visual component presented on the left. Maximum errors are shown for stimuli with the auditory component mismatched by 1dB to the right of the visual component, indicating the stimulus with perceptibly spatially corresponding components. The results of Lewald and Ehrnstein (1996) suggest that the perceived position of the auditory components of all stimuli were shifted to the left with the subjects' gaze. The IID of auditory components some distance to the right of the visual components would now be perceived as being spatially correspondent with the visual component. Data shown in figure 19 suggest that the leftward gaze shifted the perceived lateral position of the auditory components to the left by approximately 1dBIID. Similarly, the data shown in figure 20 suggest that the rightward gaze shifted

the perceived position of auditory components to the right by approximately 1dBIID.

In general, the data are suitable for their intended purpose. A difference limen for visuo-auditory spatial mismatch has been obtained. 75% detection accuracy was shown for audio-visual stimuli with a spatial non-correspondence of approximately 3dBIID. Whereas it is true that the peak in error rate is not where it would have been expected, it is also true that the portions of the discrimination functions on either side of the maxima do confirm the expected inverse relationship between audio-visual spatial mismatch and errors in detectability. The procedure tapped the ventriloquist effect, provided a systematic exploration of the influence of the auditory component, and showed that the observers were sensitive to the position of the auditory component.

Chapter 7

7.0 EFFECT OF AUDIO-VISUAL SPATIAL NON-CORRESPONDENCE ON LATERALISATION JUDGEMENTS.

The objective of the experiments to be discussed here was to assess the effects on lateralisation judgements of audio-visual spatial non-correspondence. Audio-visual spatial non-correspondence difference limen measurements - Chapter 6 - confirmed an approximately linear relationship between detectability and audio-visual spatial mismatch for stimuli with auditory components mismatched relative to visual components away from ICC. It was hypothesised that mean judgements of audio-visual stimuli with spatially non-corresponding auditory and visual components would be in the position of the visual component (c.f. Welch and Warren 1982, Radeau and Bertelson 1977, Jackson 1953). The MAH and MPH both predict a relative dominance of the visual modality in tasks where a spatial judgement of audio-visual stimuli is required independent of whether the auditory and visual components are spatially mismatched. The hypotheses indicate that relative visual dominance is a function of the visual modality's relative appropriateness and precision in tasks requiring spatial accuracy. (The MAH and MPH are described in more detail in chapter 1).

These experiments also served as an investigation of whether the relative spatial stability of the modal components of the audio-visual stimulus influenced lateralisation judgements. Subjects may have responded to the position of the auditory or visual component because it was the dominant modality in this context, or because it was relatively more consistent spatially than the other component of the audio-visual stimulus. This was assessed by comparing lateralisations of audio-visual stimuli with a relatively more stable visual component, with lateralisations of audio-visual stimuli with a relatively more stable auditory component. The MAH and MPH predict a relative dominance of the visual component rather than the auditory component in both cases (c.f. Radeau and Bertelson 1977, Welch and Warren 1982).

7.1 SUBJECTS

Six subjects took part in the experiment. All subjects had previously provided data in the audio-visual spatial difference limen measurement.

7.2 (a). Lateralisation of audio-visual stimuli: Auditory components mismatched relative to a visual component in one of three possible lateral positions.

7.2.1 STIMULI

Audio-visual stimuli with auditory and visual components varying in spatial correspondence were presented. The magnitude of spatial discrepancy between the components was varied by up to 10dBIID.

7.2.1.a Visual Components

Visual stimuli (detailed in Chapter 3) were presented in one of three possible positions on a 22cm axis drawn between the ears of a head silhouette. One of the visual positions was at intracranial center (ICC), the others were both in the left visual-field. The visual positions were analogous to auditory positions of 0dBIID (ICC), -2 dBIID and -4dBIID.

7.2.1.b Auditory Components

Auditory stimuli (detailed in Chapter 3) were presented in lateral positions spatially mismatched relative to the position of the accompanying visual component by up to -10dBIID, making a total of eleven possible audio-visual stimuli for each visual position. Tones were mismatched outwards (leftwards) from the position of the visual component, away from ICC.

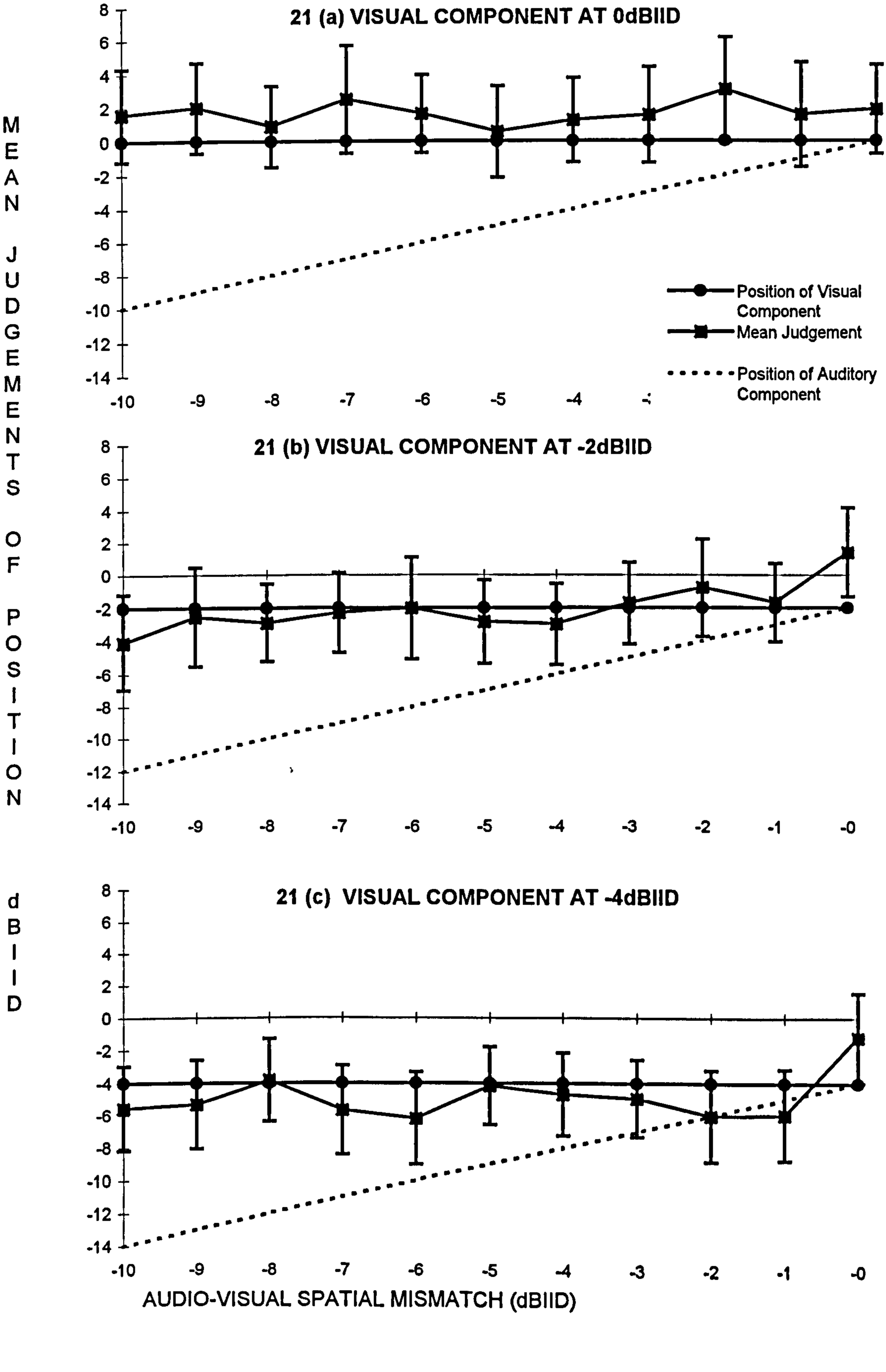
7.2.2 PROCEDURE

The procedure was as detailed in Chapter 3 Each of the eleven possible levels of audio-visual spatial mismatch was presented twenty times in each of the three visual positions. A total of 660 trials were presented in all (3x11x20). Magnitude of mismatch and visual position were selected randomly for each trial. The same instructions as those used in the experiment described in chapter 5 were given to subjects before the experiment. The possible spatial mismatch between the auditory and visual components of the audio-visual stimulus was not made explicit to subjects. Five practice trials were presented before each session.

7.2.3 RESULTS

Mean judgements of position as a function of audio-visual mismatch are plotted in figures 21a, b and c. Mean judgements were near the position of the visual component in all three visual positions tested and thus showed a relative dominance of the visual component. Figure 21(a) indicates that although subjects showed no tendency to be influenced by the position of the auditory component they did show a mild bias to respond to the right of a visual component at ICC.

FIGURE 21: Mean Judgements of Position as a Function of Audio-visual Spatial Mismatch



Measurements were repeated with the auditory component relatively more spatially stable than the visual component.

7.3 (b) Lateralisation of audio-visual stimuli: Visual components mismatched relative to an auditory component in one of three possible lateral positions.

7.3.1 STIMULI

7.3.1.a Auditory Components

Tones (detailed in Chapter 3) could be presented in one of three possible lateral positions, with IIDs of 0dBIID, -2dBIID or -4 dBIID.

7.3.1.b Visual Components

Visual stimuli (detailed in Chapter 3) were presented in lateral positions spatially mismatched relative to the position of the accompanying auditory stimulus by up to -10dbIID. A total of eleven levels of audio-visual mismatch were possible for each auditory position. Visual components were mismatched outwards from the position of the auditory component, away from ICC.

7.3.2 PROCEDURE and EQUIPMENT

The procedure and equipment were the same as those used in part (a). Each of the eleven possible levels of audio-visual spatial mismatch were presented twenty times in each of the three auditory positions. A total of 660 trials were

presented in all (3x11x20). Magnitude of mismatch and auditory position were randomised in each trial.

7.3.4 RESULTS

Mean judgements of lateral position are shown in figure 22. Mean judgements in all three panels were consistently closer to the position of the visual component than the position of the auditory component, although they tended to be biased to the right of the position of the visual component (and hence towards the position of the auditory component in most stimuli).

Mean standard deviations in judgements in part(a) and part(b) of the experiment were similar in form. A 3-way analysis of variance with relatively more stable modality (2 levels), presentation position (3 levels) and spatial mismatch (11 levels) showed no significant effects of the modality of the more stable stimuli [$F(1,5) = 0.94$, $p<0.378$], presentation position [$F(2,10) = 0.689$, $p<0.528$], or audio-visual spatial mismatch [$F(10,50) = 0.73$, $p<0.689$]. Mean standard deviations in judgements of lateral position, collapsed across all three conditions in parts (a) and (b) are shown in figure 23. There was a positive relationship between magnitude of audio-visual spatial mismatch and mean standard deviation. Whereas mean judgments remained consistent at each level of mismatch, mean standard deviations rose, indicating that the accuracy of subjects' judgements was responsive to different levels of audio-visual spatial non-correspondence.

FIGURE 22: Mean Judgements of Position as a Function of Audio-visual Spatial Mismatch. Auditory Component relatively more stable.

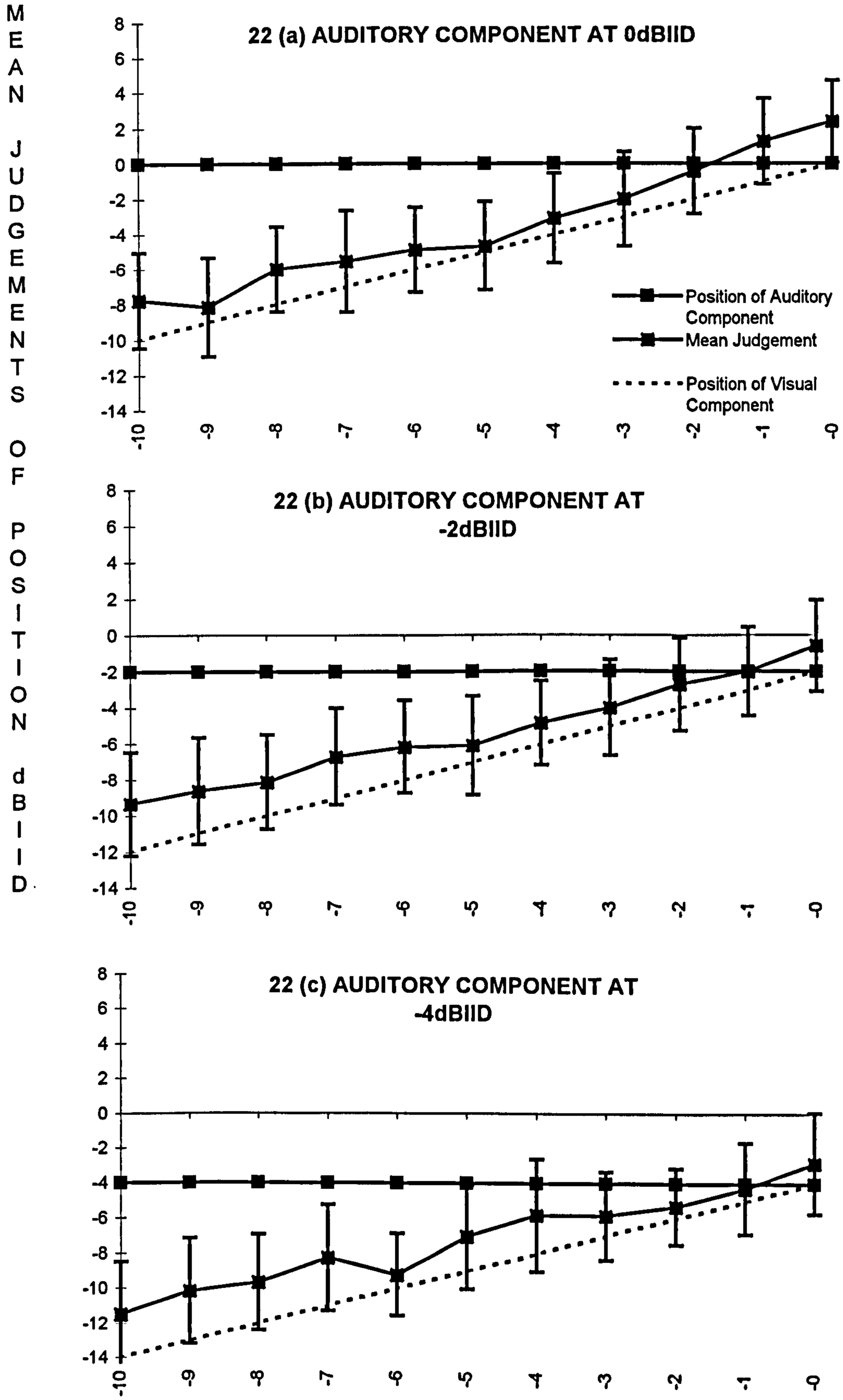
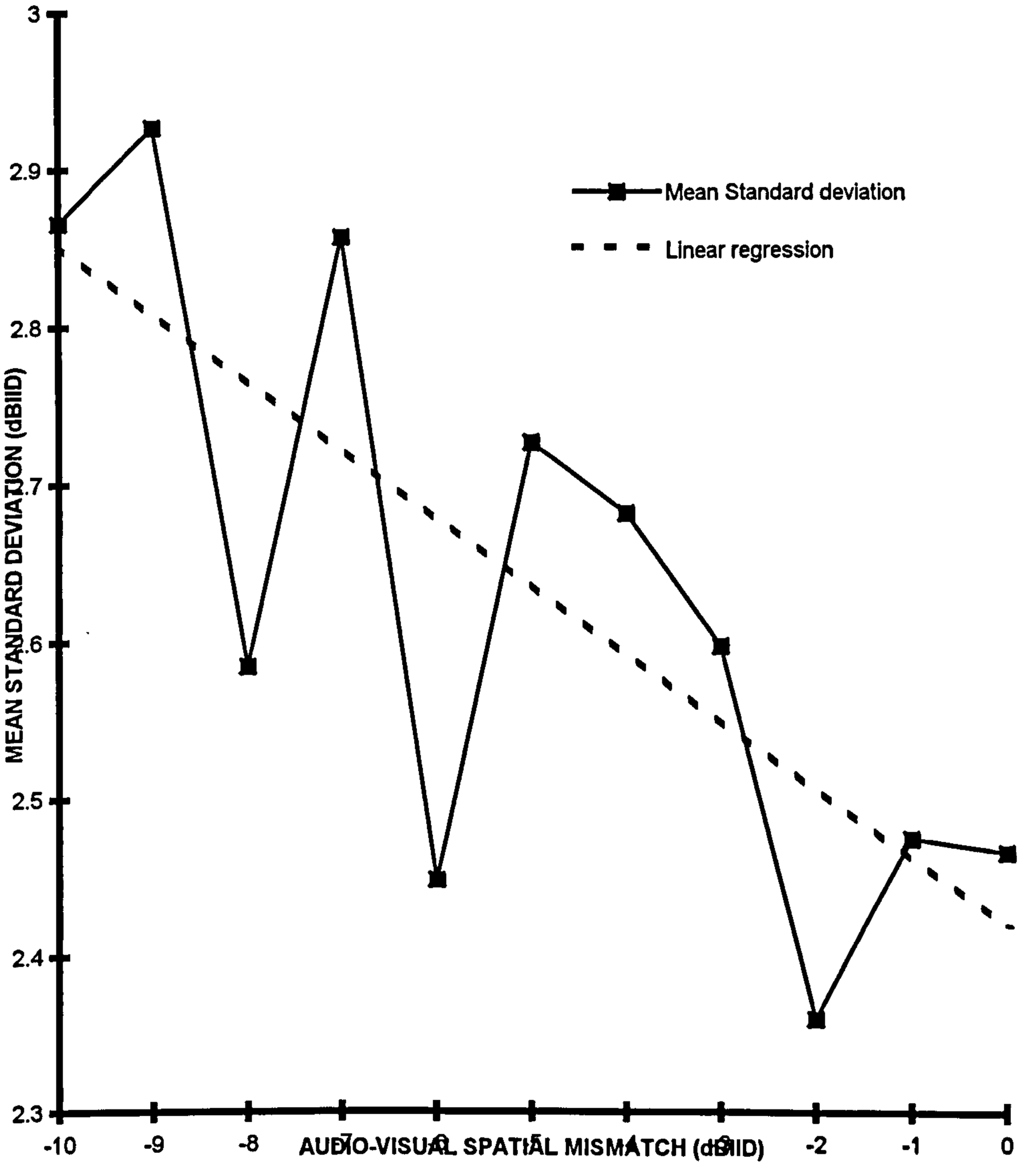


FIGURE 23: OVERALL MEAN STANDARD DEVIATIONS AT EACH LEVEL OF AUDIO-VISUAL SPATIAL MISMATCH



7.4 DISCUSSION

The results are consistent with previous studies which indicated a relative dominance of the visual modality in ventriloquism tasks (see Welch and Warren 1981). Parts (a) and (b) both showed that mean judgements of lateral position were strongly influenced by the position of the visual component irrespective of the level of audio-visual spatial mismatch. However, mean standard deviations (figure 23) increased with audio-visual spatial non-correspondence. It can be inferred from the positive relationship between mean standard deviation and audio-visual spatial mismatch that the position of the auditory component relative to the position of the visual component was relevant in this context, not simply its presence. If it were simply the presence of the auditory component that affected the variance in mean position judgements, then the magnitude of the variance should not depend on the size of the audio-visual spatial non-correspondence.

Mean judgements of the stimulus' position indicated the stimulus as being in the position of the visual component. The relative influence of the visual component in situations where the auditory and visual components of an audio-visual stimulus differ spatially has been well documented, (e.g. Jackson 1953; Welch and Warren 1981). Figures 21 and 22 show that mean judgements of the lateral position of the audio-visual stimulus were independent of the level of audio-visual spatial non-correspondence. Mean judgements shown in 21(a) show a consistent bias to the right of the visual component. It is possible that judgements presented in 21(a) effectively

represent calibration data, and as such suggest that lateralisation estimates in figures 21(b) and 21(c) should be adjusted accordingly. It is clear that if this calibration were made, mean lateralisations still show an independence of the level of audio-visual mismatch, and a relative bias of the position of the visual component. Figures 21 and 22 both show that mean judgements of lateral position were also independent of the detectability of the mismatch. Difference limen measurements made in chapter 6 showed that 75% detectability of audio-visual spatial non-correspondence was met at 3dBIID audio-visual spatial mismatch, but mean judgements of lateral position show no influence of this or any other level of mismatch detectability. This suggests that the relative dominance of the visual component in this context was not a function of a post-perceptual decision about the spatial correspondence of the auditory and visual components.

The results of the experiment described in chapter 5 indicated that variability in judgements of uni-modal stimuli was larger than variability in judgements of bi-modal audio-visual stimuli. It is possible that the increase in mean standard deviation in judgements with the level of audio-visual spatial non-correspondence in this experiment is indicative of subjects responding as if presented with a uni-modal visual stimulus rather than a bi-modal audio-visual stimulus. Using mean standard deviation as a "tag" (c.f. experiments 1, experiment 2, and Warren et al 1983) figure 23 could be interpreted as indicating that as the spatial separation between the auditory and visual components of the stimulus increased, subjects no longer based their

judgements of position on the auditory and visual components but on the visual component alone. As audio-visual spatial mismatch increased, so too did the subjects' awareness of the discrepancy (c.f. audio-visual spatial correspondence difference limen measurements made in chapter 6). As subjects became more aware of the spatial non-correspondence of the auditory and visual components their mean response accuracy became consistent with their having been presented with two stimuli in different modalities indicating different positions. The increase in mean standard deviations with the level of audio-visual spatial non-correspondence is consistent with subjects basing their judgements on uni-modal rather than audio-visual stimuli, perhaps as a function of a reduced AOU (assumption of unity) regarding the auditory and visual components of the audio-visual stimulus. By this account, increasing spatial non-correspondence weakened the evidence that the auditory and visual components of the stimulus referred to the same perceptual event.

If this interpretation of the results is correct the data are consistent with the definition of the unitary assumption provided by Welch and Warren (1981). The results indicate that the assumption of unity should be described as a continuous rather than binary assumption. It seems that, for spatial correspondence in this context at least, there is no single boundary between 'referring to the same perceptual event' and 'not referring to the same perceptual event'. Rather, there is a smooth transition between 'referring very definitely to the same perceptual event' and 'referring weakly to the same perceptual event'. This is consistent with Welch and Warren (1980), who

describe the unitary assumption as being relevant in all multi-modal situations in which components are presented in different modalities.

> "That is, situations can vary from ones in which subjects hold a very strong assumption that what they see and what they feel, for example, are actually the same physical event, to ones in which this assumption is weak or even non-existent"[1]

Difference limen measurements made in chapter 6 suggested that the detectability of audio-visual spatial mismatch was approximately linear if auditory components were mismatched relative to the visual components away from intra-cranial center - ICC. However, stimuli in part (b) of this experiment had relatively stable auditory components with visual components mismatched relative to them. Essentially, audio-visual stimuli with auditory components mismatched relative to visual components towards ICC were presented. The data shown in figure 22 suggest that the anomaly in the detectability of spatial mismatch as a function of the direction of mismatch suggested by the difference limen measurements did not influence mean judgements of the lateral position of stimuli. Judgements were strongly influenced by the position of the visual component irrespective of the level of audio-visual spatial non-correspondence.

7.5 CONCLUSIONS

The results showed that subjects indicated the perceived position of audio-visual stimuli with spatially non-corresponding auditory and visual

components as being close to the position of the visual component, irrespective of the spatial separation between the components. This is consistent with previous research which showed a similar dominance of the visual modality in tasks requiring a spatial judgement of an audio-visual stimulus with spatially non-correspondent auditory and visual components. Variability in responses increased as a function of audio-visual spatial mismatch. The data provide more evidence that the auditory component is not simply ignored in lateralisations of this kind. The data suggest that the position of the auditory component relative to the position of the visual component, and not simply its presence, affected the lateralisation judgement.

---

[1] "Immediate Perceptual Response to Intersensory Discrepancy." Welch RB & Warren DH. Page 648

Chapter 8

8.0 THE AUDIO-VISUAL TEMPORAL RELATIONSHIP.

The results of the previous experiment indicated that lateralisation judgements of the audio-visual stimuli used were strongly influenced by the position of the visual component. Similar results have been described as a visual bias, or a relative visual dominance (Welch and Warren 1981; Radeau and Bertelson 1977). Nonetheless, the results suggested that the position of the auditory component of the audio-visual stimulus relative to the position of the visual component did affect the accuracy of subjects' lateralisation judgements.

The relationship between the auditory and visual components of an audio-visual stimulus is crucial in determining whether or not the subject responds as if presented with auditory and visual stimuli, or an audio-visual stimulus. The formation of the unitary assumption - the perception that the auditory and visual stimuli refer to the same perceptual event - is affected by a number of factors. Radeau and Bertelson (1977) have said that these factors can be divided broadly into two groups: cognitive factors and structural factors. Cognitive factors are those which originate from the subjects' familiarity with the audio-visual pairing, e.g. moving lips paired with a voice. Structural factors are those which are affected by the physical nature of the auditory and visual components of the audio-visual stimulus, e.g. whether both exhibit

common changes in direction, or whether both originate from the same spatial location. Cognitive and structural factors are discussed in more detail in the section on multi-sensory perception in chapter 1.

The temporal correspondence of the auditory and visual components can be described as a structural factor. Temporal synchrony is an important cue to multi-modal integration. Subjects show a propensity to perceive asynchronous heard and seen speech as synchronous at relatively high levels of asynchrony. Minimal detectable onset asynchronies have been measured as being between 80ms (McGrath and Summerfield 1985) to 150ms (Dixon and Spitz 1980). Desynchronising the auditory and visual components by approximately 350ms has been shown to significantly reduce ventriloquism with a voice/face audio-visual pairing (Radeau and Bertelson 1977).

The objective of this experiment was to investigate whether audio-visual temporal asynchrony in spatially correspondent auditory and visual components affected subjects' lateralisation judgements of the audio-visual stimulus position. An audio-visual temporal asynchrony difference limen was measured initially. This provided data allowing the presentation of audio-visual stimuli having auditory and visual temporal differences with known detectability.

8.1 MEASUREMENT OF AN AUDIO-VISUAL TEMPORAL CORRESPONDENCE DIFFERENCE LIMEN.

8.1.1 SUBJECTS

Audio-visual temporal-correspondence difference limens of eight subjects were measured. All subjects had provided data in previous experiments. Stimulus details were as detailed in Chapter 3.

8.1.2 STIMULI

8.1.2.a Auditory Stimuli

Auditory stimuli were synthesised using the MITSYN software package (Henke 1990). 250Hz tones of 1 second in duration were presented with IID's of 0dbIID, -4dBIID or -8dBIID.

8.1.2.b Visual stimuli

Visual stimuli were 1-point bright spots presented for 1 second on an XYZ display. A silhouette mask was made to the same dimensions as those presented on the VDUs in the previous experiments. Visual stimuli were presented in lateral positions analogous to 0 dBIID, -4dBIID and -8dBIID.

After listening to pilot stimuli with a range of audio-visual asynchronies, it was evident that asynchronous audio-visual stimuli with a leading auditory component were difficult to identify as asynchronous even at relatively large asynchronies. Asynchronous audio-visual stimuli with a lagging auditory

component were perceptibly asynchronous at substantially smaller asynchronies. The reason for this asymmetry is unclear. It may be that the offset of the stimulus is an important cue to audio-visual asynchrony, and an auditory lag is easier to detect than a visual lag. For the purposes of this experiment, stimuli with leading visual components were used throughout, because asynchronies were easier to identify and smaller gradations of asynchrony could be assessed. Since both components were 1 second in duration, the offset of the visual component was always prior to the offset of the auditory component. Eleven levels of audio-visual asynchrony were presented, ranging from 25ms to 275ms in 25ms steps. The asynchrony was calibrated by comparing the relative onsets of the auditory and visual components on a two-trace oscilloscope. A photocell attached to the visual display provided a signal at the onset of the visual stimulus.

8.1.3 EQUIPMENT

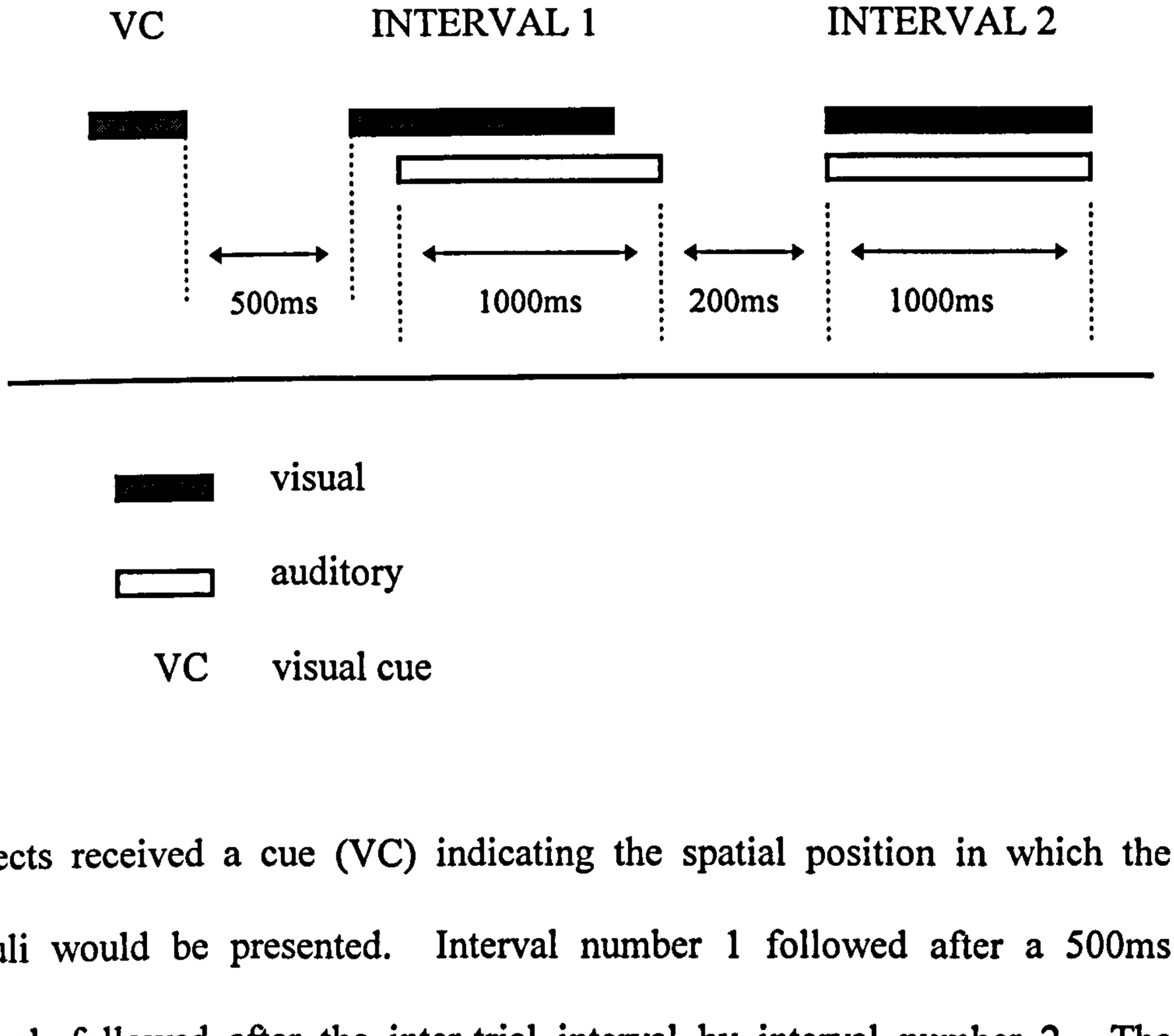
Experimental equipment was as detailed in the general equipment section, except for the VDU screen. The visual components of the audio-visual stimuli were presented between the ears of a head silhouette mask mounted in front of a Hewlett Packard 1304A, 12-inch XYZ display.

8.1.4 PROCEDURE

A 2I-2AFC procedure was used. Subjects were presented with two audio-visual stimuli with an ISI of 200ms (figure 24). In each trial, one of the intervals contained a stimulus with synchronous components, the other

stimulus had components which onset and offset asynchronously, as outlined above.

Figure 24



Subjects received a cue (VC) indicating the spatial position in which the stimuli would be presented. Interval number 1 followed after a 500ms interval, followed after the inter-trial interval by interval number 2. The subjects' task was to identify the interval in which the auditory and visual components were *asynchronous* by pushing one of a pair of keys marked 1 and 2.

Eleven levels of audio-visual asynchrony were presented 10 times each in the three spatial locations in random order. Spatial presentation position and mismatch level were chosen randomly on each trial. Each of the eleven levels

of asynchrony was presented once in a practice session. Subjects were provided with feedback in the practice trials. A correct response was indicated to the subject by a central flashing dot.

8.1.5 RESULTS

A positive relationship between detectability and size of audio-visual asynchrony was found for asynchronies greater than 50ms. A 2 -way analysis of variance with audio-visual asynchrony (11 levels) and presentation position (3 levels) showed no significant main effect of the position of stimulus presentation. [$F(2,14)=0.57$, $p=0.579$]. The main effect of asynchrony was shown to be significant [$F(10,70)=62.98$, $p<0.001$]. The interaction between the two factors was not significant [$20,140)=0.46$, $p=0.977$]. Individual results collapsed across presentation positions are shown in figure 25. Mean errors collapsed across all subjects are shown in figure 26.

FIGURE 25: MEAN PERCENT CORRECT AS A FUNCTION OF AUDIO-VISUAL ASYNCHRONY. INDIVIDUAL DATA.

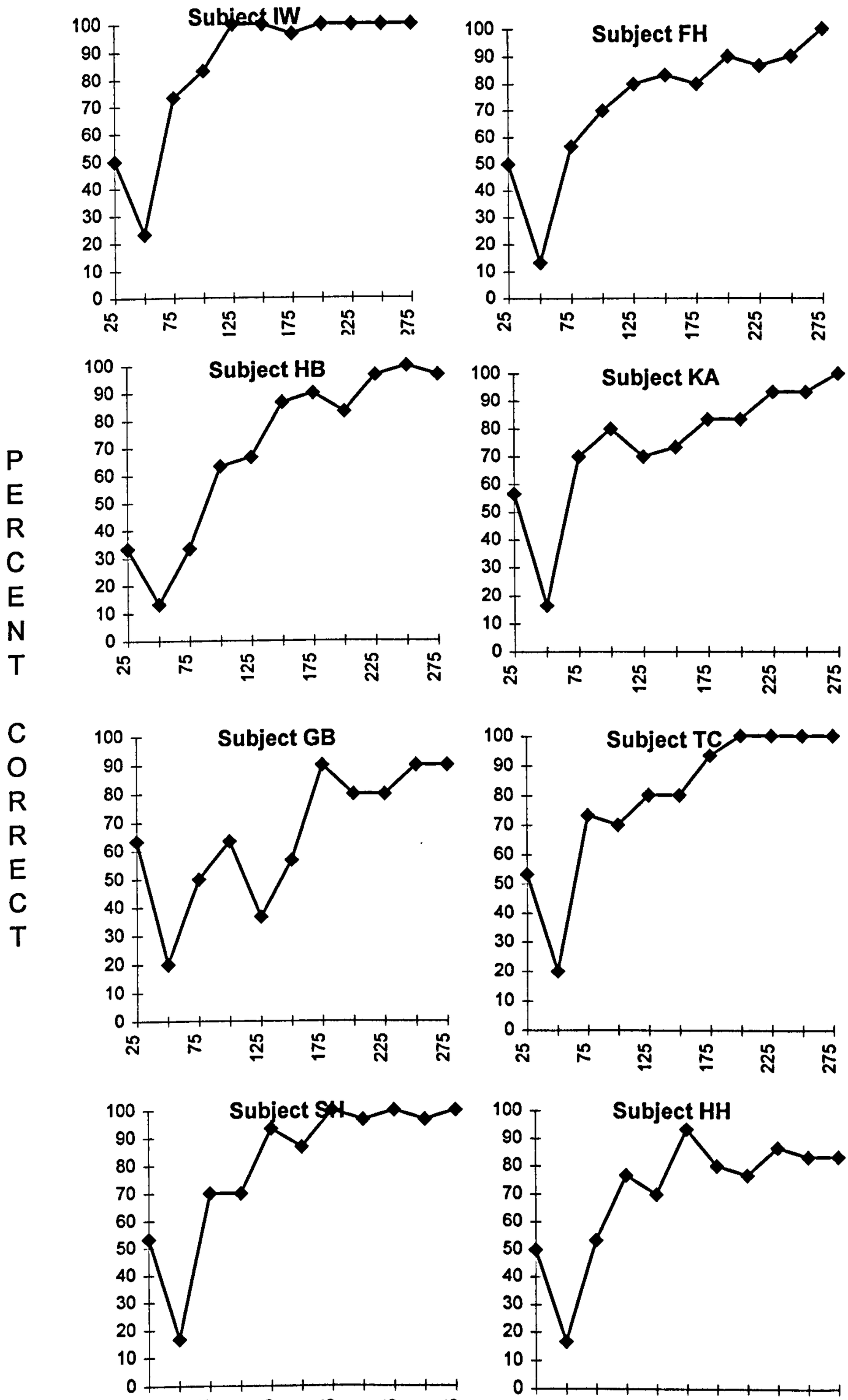
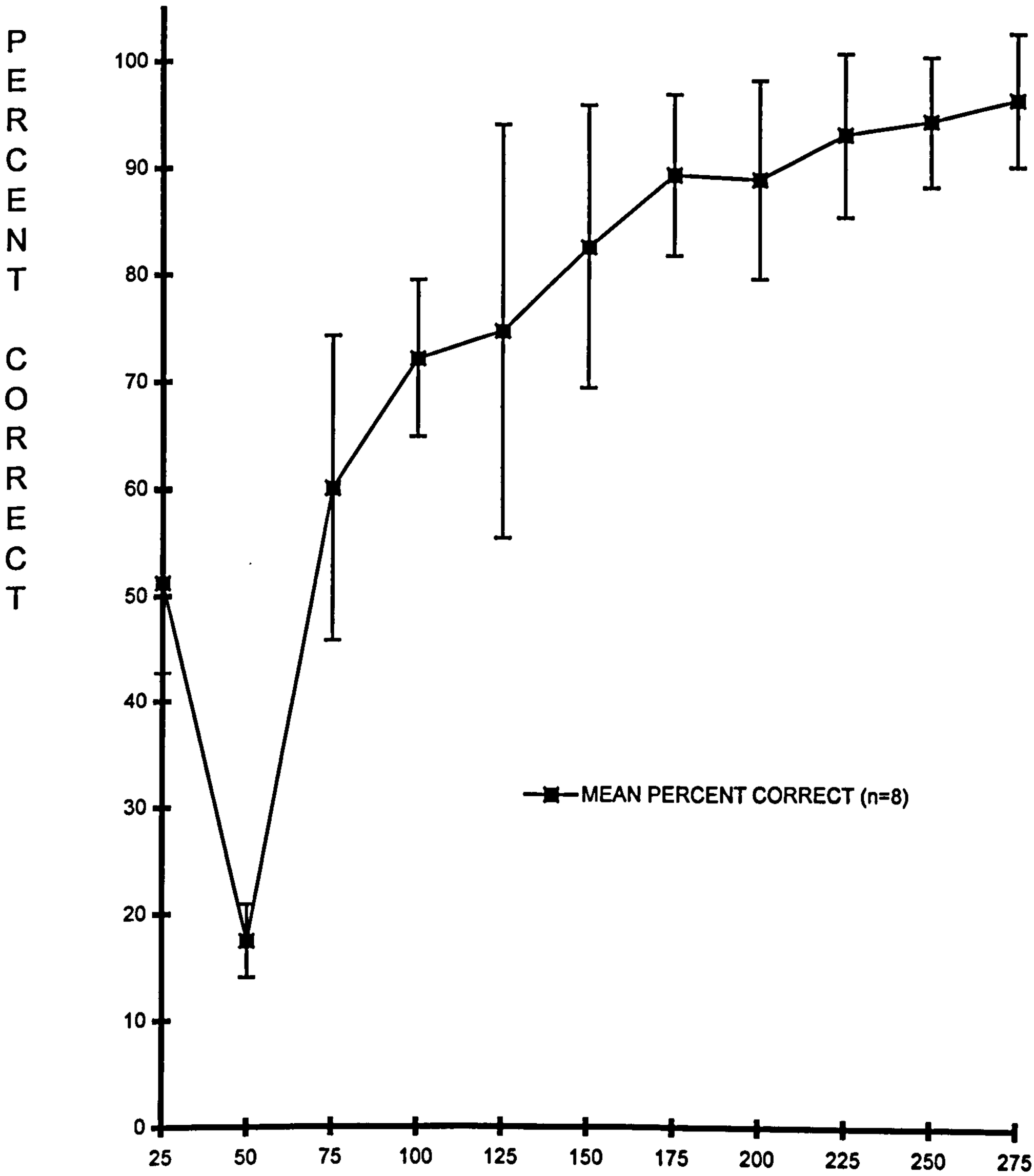


FIGURE 26: MEAN PERCENT CORRECT AS A FUNCTION OF AUDIO-VISUAL ASYNCHRONY, SHOWING SUBJECT DEVIATION.



8.1.6 DISCUSSION

The results indicated that ability to detect temporal asynchronies in the audio-visual stimuli presented improved with the asynchrony for temporal mismatches above 50ms. At 50ms asynchrony, performance on the task was markedly below chance. This anomaly can be restated as a propensity to choose the temporally matched pair over the temporally mismatched pair when the audio-visual mismatch was 50ms.

The tendency for subjects to choose the 'incorrect' interval at 50ms asynchronies may be attributable to differences in the detection latency for auditory and visual stimuli. Response times to simple visual stimuli have been measured as being approximately 40-50ms longer than response times to auditory stimuli (Niemi & Naatanen 1981; Elliot 1968; Rutschmann & Link 1964; Poppel 1988, Lewkowicz 1996). This suggests that the time between stimulus presentation and perceptual impact on the observer is longer for visual stimuli than for auditory stimuli. Poppel (1988) proposed a hypothetical 'horizon of simultaneity' which is approximately ten meters from the subject. His measurements of reaction times to auditory and visual stimuli suggested a visual lag of approximately 40ms - the time taken for sound to travel approximately 10 meters. He suggested that light and sound leaving a point ten meters away from the observer will arrive at their 'central' neural destination simultaneously. The actual identity of the 'central' position is unclear. Neural evidence presented in chapter 1 suggests that a 'central'

position for spatio-temporally corresponding auditory and visual components might be the superior colliculus, but Poppel is not clear about where, or what exactly the central position might be. However, it follows that the visual components of audio-visual stimuli with physically synchronous auditory and visual components would be perceived as lagging behind auditory components if the audio-visual source is less that 10 meters from the subject. If we make the assumption that visual lag is approximately 50ms, the nominal audio-visual asynchronies in stimuli presented to subjects in this experiment did not provide the intended asynchronies.

When presented with a trial with the configuration shown in figure 27, the correct response would be to choose interval number 1, the interval in which the auditory and visual components are asynchronous. Stimuli presented in interval number 2 are physically synchronous (27a). When the 50ms visual lag is taken into consideration (27b), the task is no longer one in which they must identify the interval with asynchronous components, instead subjects must identify the interval in which the auditory and visual components are 'more asynchronous'. The asynchrony in the target interval (interval 1) is reduced from 275ms to 225ms. The previously synchronous components in interval number 2 are now asynchronous, the visual component lagging behind the auditory component by 50ms.

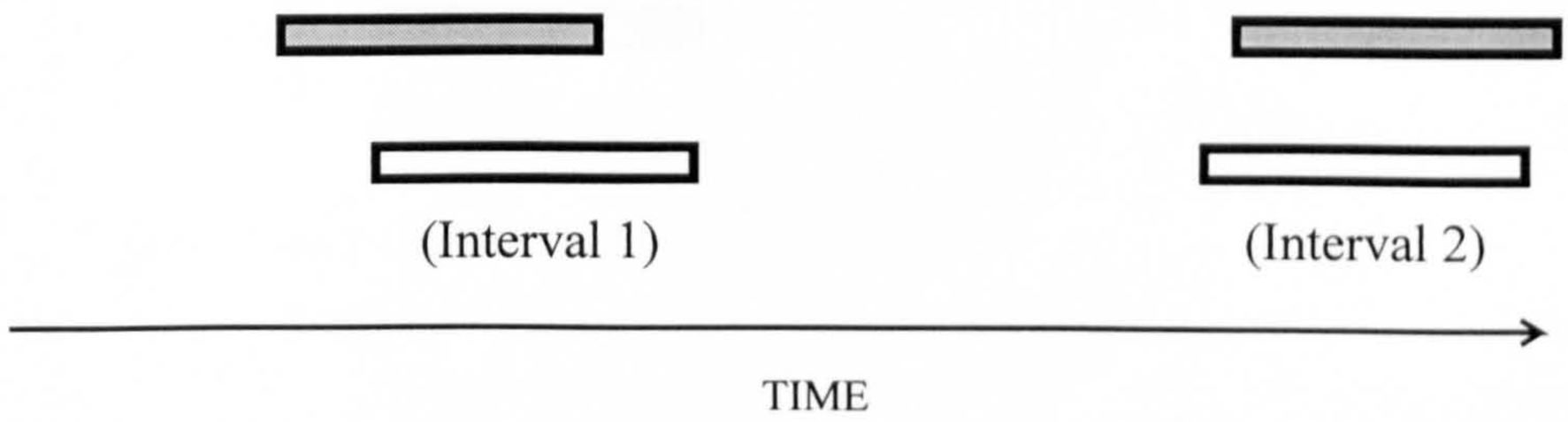
FIGURE 27

Target stimuli with audio-visual asynchrony of 275ms

VISUAL AUDITORY

(27a) Physical representation

(27b) Perceptual representation



Mean responses (figure 25) showed that subjects responded correctly by indicating interval number 1 on the majority of trials with this configuration. This was presumably because of the considerable temporal difference in the relative asynchronies in the two intervals.

FIGURE 28

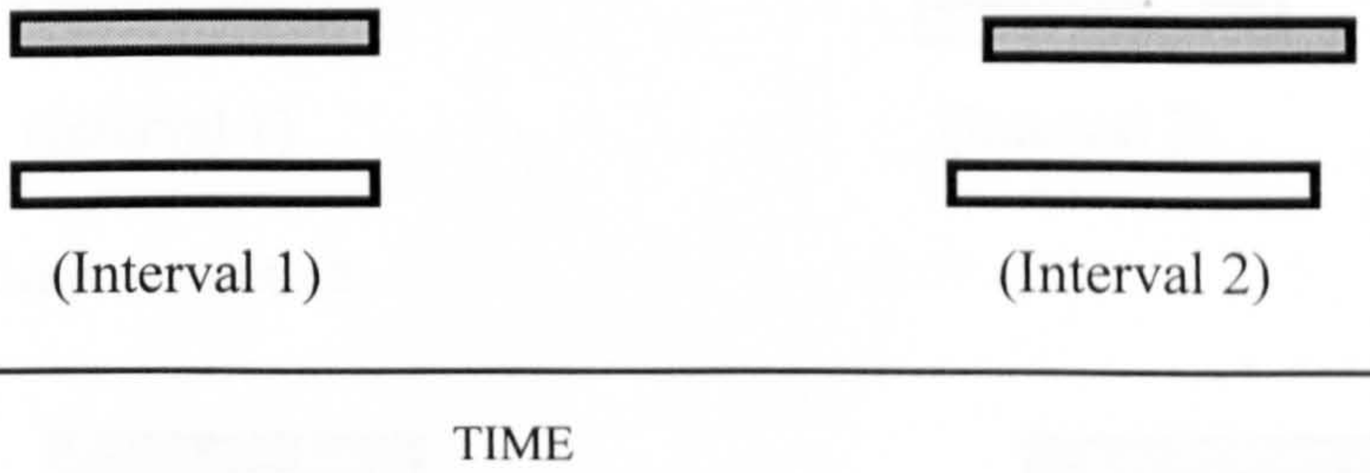
Target stimuli with audio-visual asynchrony of 50ms

VISUAL AUDITORY

(28a) Physical representation

(Interval 1) (Interval 2)

(28b) Perceptual representation



(Interval 1) (Interval 2)

TIME

Applying the same reasoning to the trial configuration shown in figure 28 suggests a possible explanation for the subjects' consistent choice of the wrong interval in trials in which the target stimulus had an asynchrony of 50ms. When the 50ms visual lag is taken into consideration (28b) the auditory and visual components of the stimulus in interval number 2 are perceived as asynchronous, the visual component lagging behind the auditory component by 50ms. The auditory and visual components in interval number 1 now appear to be synchronous. The subject responds with interval two, the interval

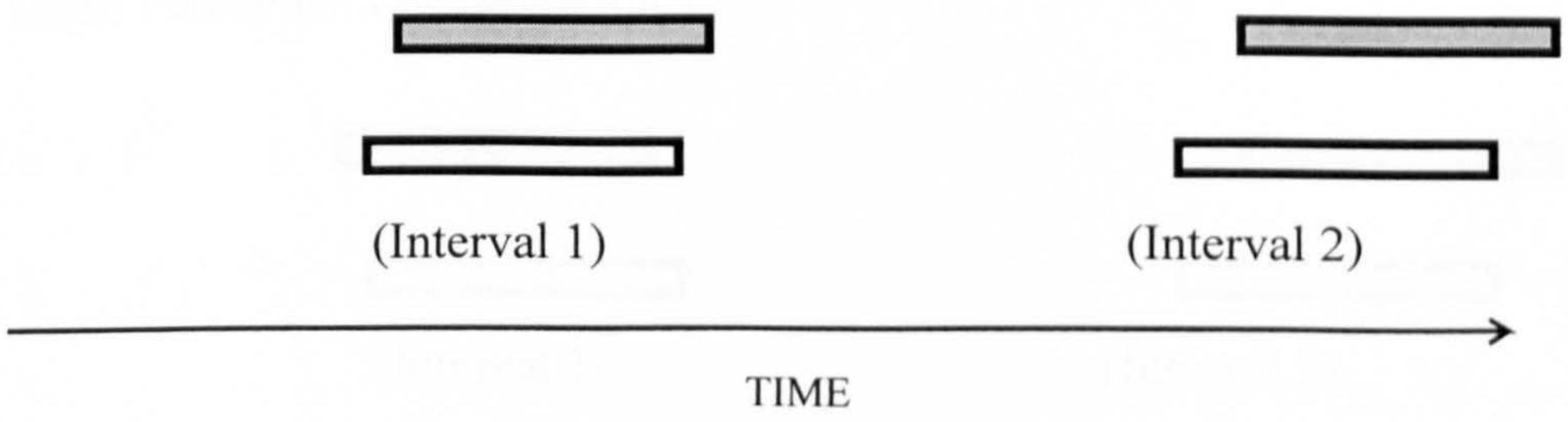
with perceptually asynchronous components. Mean performance, shown in figure 26, indicated that subjects responded incorrectly on approximately 82% of trials with a 50ms asynchrony.

FIGURE 29

Target stimuli with audio-visual asynchrony of 25ms

(29a) Physical representation

(29b) Perceptual representation



When the 50 ms visual lag is taken into consideration in trials with targets with components asynchronous by 25ms, subjects must choose between the intervals shown in figure 29b. Both intervals have asynchronous components, with a difference of 25ms between the asynchronies. Subjects guess at the correct pairing. This is consistent with the mean performance levels shown in

figure 26, which show that subjects performance on trials with asynchronies of 25ms was at chance.

Mean performance on trials with asynchronies of 75ms or 100ms (figure 30) was better than chance.

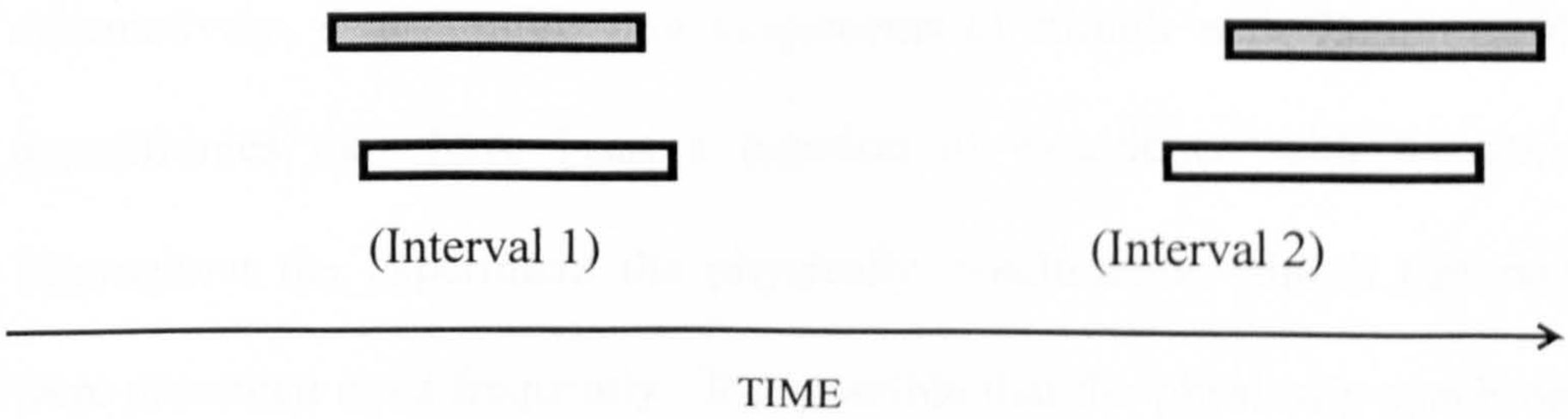
FIGURE 30

Target stimuli with audio-visual asynchrony of 75ms

VISUAL AUDITORY

(30a) Physical representation

(30b) Perceptual representation



When the 50ms visual lag is taken into consideration (30b) the subjects' choice was between two asynchronous intervals, and as such performance might be expected to be at chance level (figure 30). In interval number 1, the visual component precedes the auditory component by 25ms. In interval number 2 the visual component lags behind the auditory component by 50ms.

The difference in the sizes of the asynchronies is 25ms, as it was in the example shown in figure 29. However, in this example the order in which the auditory and visual components onset and offset is different in each interval. In pilot listening trials, asynchronies in audio-visual stimuli with the auditory component leading the visual component were considered harder to detect than asynchronies in which the visual component led the auditory component, although the reason for this is unclear. In trials in which asynchronies of 75ms or 100ms were presented, the 50ms visual lag meant that subjects were forced to choose between two audio-visual stimuli with asynchronous components. If they found stimuli with the auditory component leading the visual component harder to detect as asynchronous, subjects would be likely to choose stimuli in interval number 1 as having asynchronous components on more occasions than they chose stimuli in interval number 2.

Alternatively, it is possible that judgements of stimuli with 75ms or 100ms asynchronies may have been a function of experience with the stimuli. Throughout the experiment the physically synchronous stimuli (interval 2) were presented most frequently. It is possible that the physically synchronous audio-visual pairing may form a background, or template against which stimuli are judged. In figure 30, alternative number 2 (30b) fits the template, and as such is rejected as the correct interval.

The discussion of the results has so far been concerned with the onset asynchrony of the auditory and visual components of the audio-visual stimuli.

It is clear, however, that because components of equal duration were used, any onset asynchrony was combined with an offset asynchrony of equal duration. It may have been that subjects were attending to the offset of the stimuli for their cue to the relative synchrony of the two alternatives. If visual lag is taken into consideration a similar explanation of the results can be offered.

8.1.7 CONCLUSION

The results have shown that the identification of temporal asynchrony in the components of audio-visual stimuli improved with the magnitude of the temporal non-correspondence for asynchronies greater that 50ms. The results are consistent with the notion of a visual lag of approximately 50ms (c.f. Poppel 1988). On average, 75% correct performance was achieved with asynchronies of 125ms.

Chapter 9

9.0 EFFECT OF AUDIO-VISUAL TEMPORAL NON-CORRESPONDENCE ON LATERALISATION JUDGEMENTS.

Temporal difference limen measurements showed that the mean detectability of audio-visual asynchrony improved with the magnitude of temporal mismatch in the auditory and visual components of the audio-visual stimuli if the asynchrony was greater than 50ms. The objective of this experiment was to investigate the effect of desynchronising auditory and visual stimulus components on lateralisation judgements of audio-visual stimuli.

In this experiment both the auditory and visual components of the stimulus were presented in analogous spatial positions, and as such, no effect of temporally desynchronising the components was expected on mean judgements of position.

Mean standard deviations in judgements of spatially corresponding uni-modal stimuli were larger than mean standard deviations in judgements of spatially corresponding bi-modal stimuli (c.f. chapter 5). As temporal asynchrony increased audio-visual structural correspondence should be weakened, and as a consequence of this it is likely that subjects would form a weakened AOU. It was hypothesised, therefore, that mean judgement accuracy, measured in mean

standard deviations, would decrease as a function of increasing audio-visual asynchrony.

9.1 SUBJECTS

Six subjects took part in the experiment. All subjects had previously provided data in the audio-visual temporal non-correspondence difference limen measurement.

9.2 EQUIPMENT

The equipment was the same as that used in the audio-visual temporal non-correspondence measurements.

9.3 STIMULI

Audio-visual stimuli were presented with auditory and visual components varying in temporal correspondence.

9.3.1 Auditory stimuli

Tones (detailed in Chapter 3) were presented in one of three lateral positions with IID's of 0db, -4dB or -8dB.

9.3.2 Visual stimuli.

Visual stimuli were 1-point bright spots presented for 1 second on the XYZ display. Visual stimuli were presented within the head silhouette in lateral positions analogous to 0 dBIID, -4dBIID and -8dBIID.

9.4 AUDIO-VISUAL TEMPORAL RELATIONSHIP

Four conditions were presented. Three conditions were presented in which the auditory and visual components of the audio-visual stimuli varied in temporal asynchrony. In condition 1 the auditory and visual components were synchronous. In condition 2 the visual component preceded the auditory component by 125ms - Mean 75% correct temporal difference limen. In condition 3 the visual component preceded the auditory component by 275ms. In condition 4 uni-modal visual stimuli were presented.

9.5 PROCEDURE

Subjects were presented with a visual cue, followed by the stimulus (c.f. figure 14 - chapter 3). Subjects were required to adjust the lateral position of an auditory pointer until it matched that of the stimulus. No time limit was put on the matching process. Conditions were blocked and the order of presentation of conditions 1 – 3 was counterbalanced across subjects. Condition 4 was presented in a random position in the condition sequence. Stimuli were presented in the three lateral positions twenty times each per condition in random order. Rest intervals were allowed after each condition.

9.6 RESULTS

Mean judgements of lateral position as a function of stimulus position are shown in figure 31. Although mean judgements in all conditions reflect the position of the stimulus, a bias to respond to the left of the stimulus is shown. A 2-way analysis of variance with condition (4 levels) and stimulus position (3 levels) as factors showed no significant effect of condition [$F(3,15)=0.12$, $p<0.946$]. Stimulus position was shown to be a significant factor in the analysis [$F(2,10)=254.19$, $p<0.001$]. The interaction between the two factors was not significant [$F(6,30)=0.81$, $p<0.572$].

Figure 32 shows mean standard deviations in subjects' judgements of the lateral position of stimuli. An influence of stimulus condition is suggested by the vertical separation between the mean standard deviation function of condition 1 and the mean standard deviation functions of conditions 2, 3 and 4. A 2-way analysis of variance with condition (4 levels) and stimulus position (3 levels) showed no significant effect of stimulus position [$F(2,10)=1.29$, p,0.318]. The effect of condition approached significance [$F(3,15)=2.65$, $p<0.087$], consistent with the observation made earlier. The interaction between the two factors was not significant [$F(6,30)=0.26$, $p<0.952$].

FIGURE 31: MEAN JUDGEMENTS OF LATERAL POSITIONS OF STIMULI WITH VARING LEVELS OF AUDIO-VISUAL TEMPORAL CORRESPONDENCE

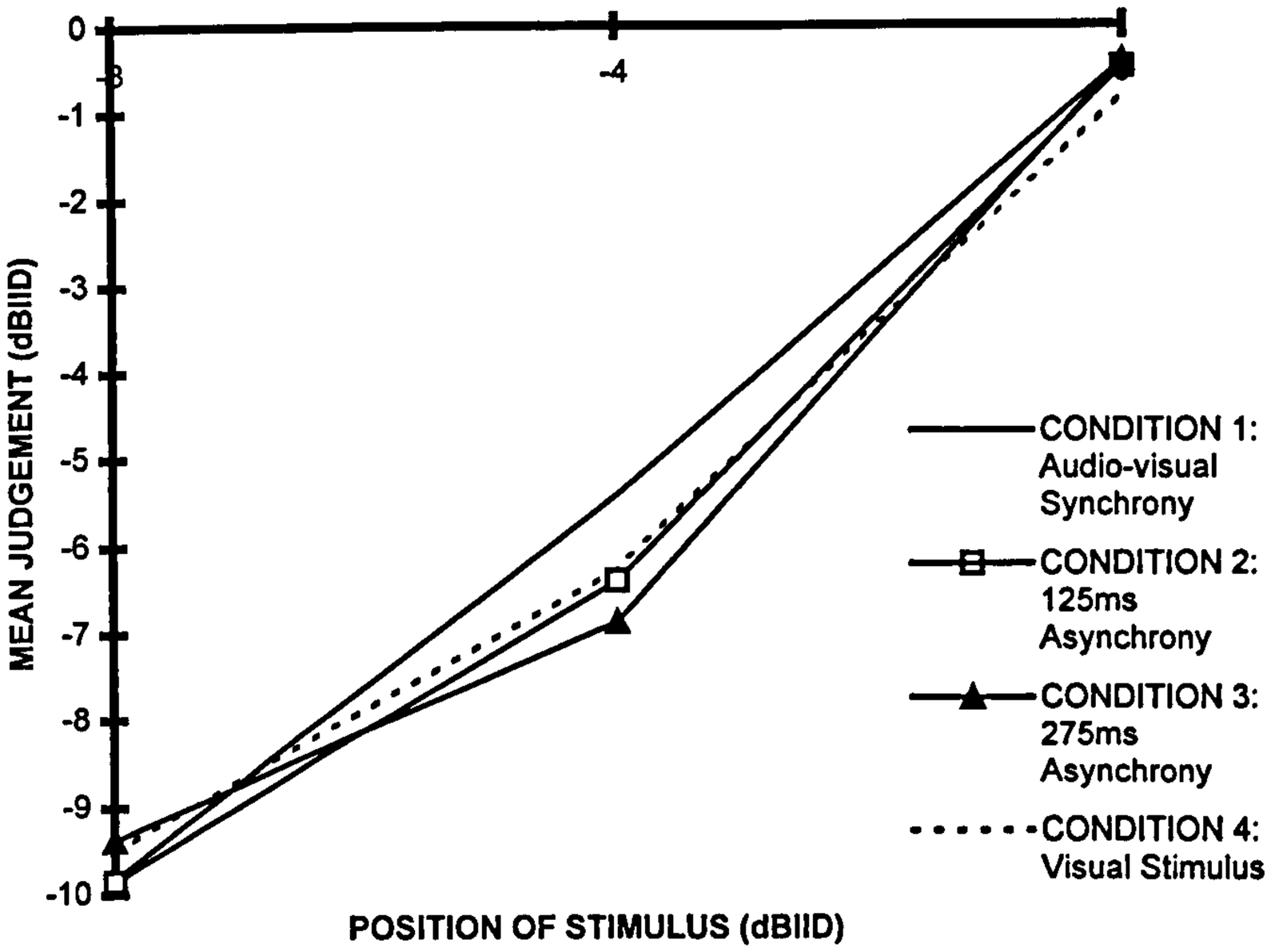
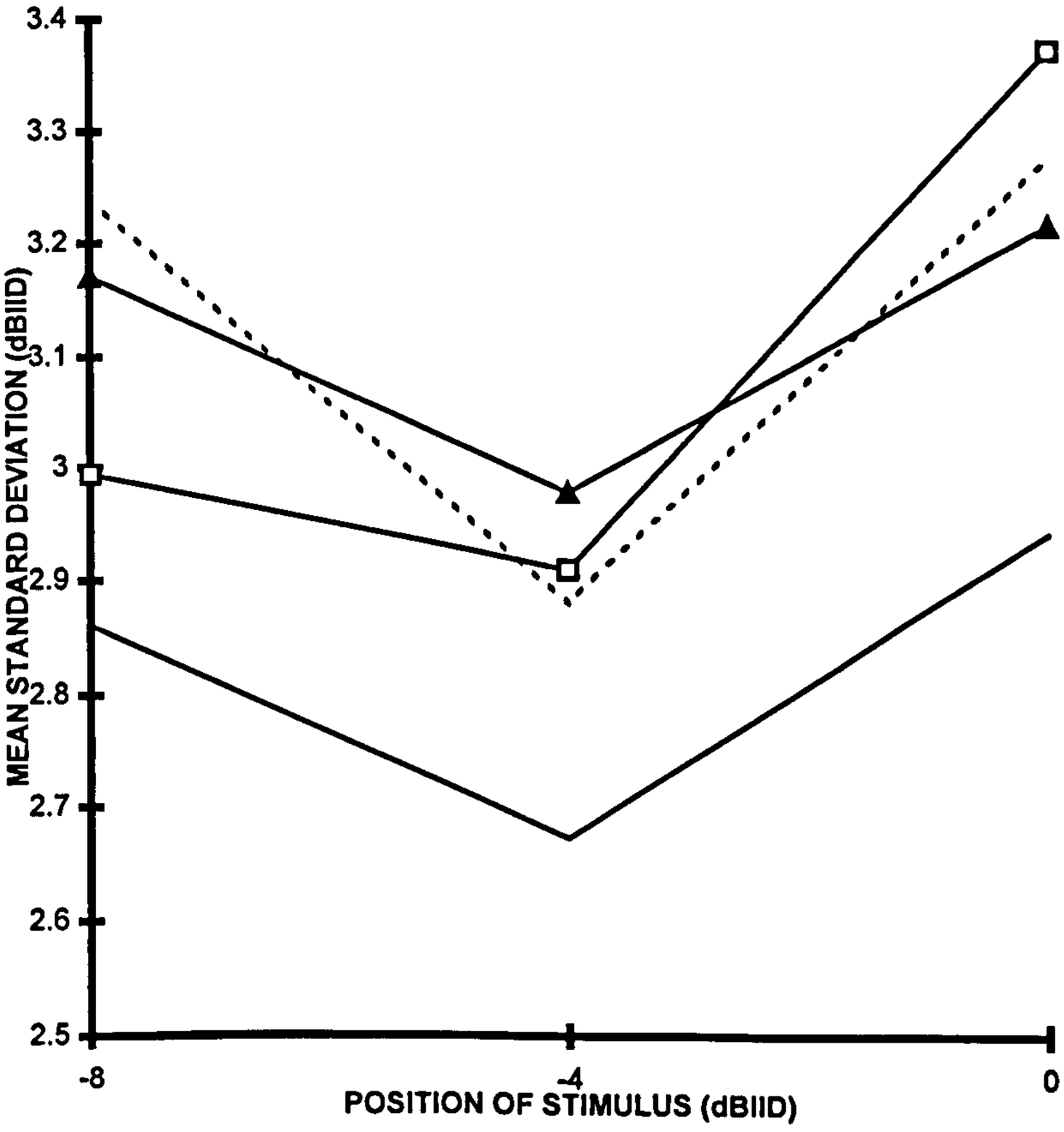


FIGURE 32: MEAN STANDARD DEVIATIONS



9.7 DISCUSSION

Mean lateralisation judgements (figure 31) were independent of stimulus condition, as expected. This was consistent in part with the results of the experiment described in chapter 5, which showed than mean judgements of audio-visual stimuli with spatially corresponding modal components (c.f. stimuli in condition 1) were not significantly different from mean judgements of uni-modal visual stimuli (c.f. stimuli in condition 4). The results were also consistent with those of the previous experiment, in which the spatial structural variable rather than the temporal structural variable was manipulated.

Although mean lateralisation judgements in chapter 5 were independent of the condition in which the stimuli were presented, variability in judgements, expressed as mean standard deviations, was significantly lower when stimuli were presented audio-visually, than when stimuli were presented uni-modally. Data in this experiment exhibited a similar trend (figure 32), with differences in variability in judgements as a function of stimulus condition approaching significance ($p<0.087$). As hypothesised, the data (figure 32) suggested that the accuracy of judgements of synchronous audio-visual stimuli (condition 1) was greater than the accuracy of judgements of audio-visual stimuli with asynchronous components (conditions 2 and 3), and of judgements of uni-modal visual stimuli (condition 4).

The data suggest that mean judgements of audio-visual stimuli with temporally non-correspondent auditory and visual components were characteristic of subjects having been presented with uni-modal stimuli (c.f. chapter 5). Temporally dissociating the auditory and visual components had no influence on mean lateralisation judgements of the stimuli but the variance in lateralisation judgements - judgement accuracy - was decreased. Similar results were shown in the previous experiment, where spatially separating the auditory and visual components of audio-visual stimuli increased variance in lateralisation judgements, but did not affect subjects mean judgement of position. Bregman (1990) notes numerous cases in which temporal relationships and spatial location have been shown to affect the parsing of the auditory scene. The audio-visual scene can be described similarly. Welch and Warren (1980) identify cross-modal spatial and temporal relationships as structural factors in the formation of the unitary assumption. Plausibly, temporally or spatially separating the auditory and visual components of an audio-visual stimulus reduces the strength of the assumption of unity regarding the components, and the weakend AOU is reflected in the data of figure 32.

The influence of modal asynchrony is evident at relatively low levels of asynchrony, c.f. condition 2 in which the magnitude of asynchrony used (125 ms) corresponded to the 75% correct point on the psychometric function for detection of asynchrony (Chapter 8). Since only stimuli in one condition were presented in each block, it is possible that the repeated exposure to stimuli

with 125 ms asynchrony in condition 2 allowed subjects to listen more analytically, which meant that this asynchrony was more detectable than the data of Chapter 8 would suggest. The slope of the psychometric function at 75% detectability is relatively steep (figure 26 - Chapter 8), and a small shift in the function facilitated by such analytical listening would result in a relatively large change in the detectability of a 125ms audio-visual asynchrony. This is consistent with the data, since mean variance in judgements of stimuli with a 125ms component asynchrony (condition 2) was closer to mean variance in judgements of uni-modal visual stimuli (condition 4) and stimuli with a 275ms asynchrony (condition 3) than mean variance in judgements of synchronous audio-visual stimuli (condition 1), suggesting that subjects were sensitive to the 125ms asynchrony.

The data indicate that the auditory component, while not dominant in lateralisation tasks (c.f. Jackson 1953; Welch and Warren 1982), is not ignored. Moreover, the temporal relationship between the auditory and visual components of an audio-visual stimulus was a factor in the relative accuracy of judgements of the position of the audio-visual stimulus.

9.8 CONCLUSION

The temporal relationship between the auditory and visual components of the audio-visual stimulus was a factor in the accuracy of lateralisation judgements. Although the position of one component may have been relatively dominant,

both components played a role in the lateralisation task. The mean lateralisation judgement data (figure 31) and the mean accuracy data (figure 32) are consistent with the mean lateralisation judgement and mean accuracy data in the previous experiment. The similarity in the two sets of data suggests that the temporal and spatial correspondence of the modal components may play equivalent roles in the perception of audio-visual correspondence.

Chapter 10

10.0 THE AUDIO-VISUAL SPATIO-TEMPORAL RELATIONSHIP.

The previous experiment investigated judgements of the lateral positions of audio-visual stimuli as a function of audio-visual asynchrony. The experiment described in chapter 7 looked at how lateralisation judgements were affected by spatially separating the auditory and visual components of the stimulus. The results of both experiments indicated that manipulating the spatial or temporal correspondence of the auditory and visual components of the audio-visual stimulus could affect lateralisation judgements. Mean responses remained in the position of the visual component in both cases, irrespective of the size of the audio-visual temporal or spatial mismatch. Variance in response was positively related to the size of mismatch. This suggested some sensitivity to temporal and structural audio-visual factors when they were manipulated individually.

It was the objective of this experiment to investigate the effects on lateralisation responses of simultaneously varying the temporal and spatial correspondence of the modal components.

The results of previous experiments indicated that the position of the visual component would dominate responses if the components of the stimuli were

mismatched spatially or temporally. In each experiment, one structural variable corresponded auditorily and visually in each stimulus presentation, providing subjects with predictability and historical evidence which facilitated the unitary assumption. In this experiment both temporal and spatial factors were varied in each stimulus presentation. Subjects were unable to predict any consistency in the stimulus, and reliable historical evidence about the stimulus was not available for the formation or strengthening of the unitary assumption. This is likely to have led to an at best weak, and at worst non-existent unitary assumption. The experiment explored how manipulation of spatial and temporal structural variables would affect lateralisation responses in the context of a weakened assumption of unity.

One factor which may affect the accuracy of lateralisation judgements is stimulus unpredictability. Simultaneously manipulating auditory and visual temporal and spatial correspondence increases stimulus unpredictability. Unpredictability can be described in terms of the so called 'historical factors' (Welch and Warren 1980) that influence the strength of the unitary assumption (chapter 1). Any spatial or temporal consistency in the audio-visual stimulus provides the perceiver with historical evidence in favour of a unitary assumption on each stimulus presentation. If the modal components in the multi-modal stimulus seldom or never correspond spatially and/or temporally there is a lack of historical evidence that the two components refer to the same perceptual event, which weakens the unitary assumption. This suggests that simultaneously manipulating the structural variables would lead to a weaker

unitary assumption than independent manipulation of the spatial or temporal correspondence, as in the previous experiments.

In experiments discussed previously (chapters 7 and 9), an increase in response variance as a function of audio-visual spatial or temporal non-correspondence was found, possibly due a weakened AOU, itself a function of audio-visual structural correspondence. In this experiment, the unitary assumption is likely to be weaker at all levels of audio-visual non correspondence than it was in the experiments described in chapters 7 and 9, due to stimulus unpredictability, and a consequent lack of 'historical' factors available for the strengthening of the AOU. By this account accuracy, measured in terms of the variance in responses in these experiments, was expected to be relatively low, and more similar at all levels of audio-visual mismatch than at different levels of audio-visual spatial mismatch presented in chapters 7 and 9.

In summary, the results of previous experiments indicated that the position of the visual component would dominate responses if the components of the stimuli were mismatched spatially or temporally. In each experiment, one structural variable corresponded auditorily and visually in each stimulus presentation, providing subjects with predictability and historical evidence which facilitated the unitary assumption. In this experiment both temporal and spatial factors were varied in each stimulus presentation. Subjects were unable to predict any consistency in the stimulus, and reliable historical evidence

about the stimulus was not available for the formation or strengthening of the unitary assumption. This is likely to have led to an at best weak, and at worst non-existent unitary assumption. The experiment described here explored how manipulation of spatial and temporal structural variables would affect lateralisation responses in the context of a weakened assumption of unity.

10.1 SUBJECTS

Six subjects with thresholds within the normal range took part in the experiment. All subjects had provided data in the previous experiment.

10.2 EQUIPMENT

The equipment was the same as that used in the previous experiment.

10.3 STIMULI

Stimuli were as detailed in Chapter 3. Audio-visual stimuli were presented with auditory and visual components varying in temporal correspondence, spatial correspondence and lateral presentation position.

10.3.1 Auditory stimuli

1 second, 250Hz tones were presented in different lateral positions given by the IID.

10.3.2 Visual stimuli.

Visual stimuli were 1-point bright spots presented for 1 second on the XYZ display. Visual stimuli were presented within the head silhouette in lateral positions analogous to 0 dBIID, -2dBIID and -4dBIID. These are referred to from here on as 'visual presentation positions'.

10.4 AUDIO-VISUAL TEMPORAL RELATIONSHIP

The auditory and visual components of the audio-visual stimuli varied in temporal asynchrony. Audio-visual stimuli with one of three levels of asynchrony were presented. Modal components could be synchronous, or the visual component could precede the auditory component by 125ms (mean 75% correct temporal difference limen measured in experiment 7) or 275ms. Temporal asynchronies were calibrated as described in the previously detailed temporal difference limen measurements.

10.5 AUDIO-VISUAL SPATIAL RELATIONSHIP

Auditory stimuli were presented in 3 different lateralisations relative to the position of the visual components, making a total of nine possible presentation positions. Tones were presented in the same position as the visual stimulus (0dBIID difference), -4dBIID or -10dBIID relative to the position of the visual component. The auditory component was always mismatched to the left of the visual component, away from intra-cranial center (ICC).

10.6 INSTRUCTIONS

Subjects were informed that they were to be presented with audio-visual stimuli. Subjects were all familiar with the general procedure as they had been subjects in the previous experiment. They were not informed of how the stimuli varied, or of the accuracy of their performance at any time during the experiment.

10.7 PROCEDURE

The procedure was similar to that used in chapter 5. Subjects were presented with a visual cue, followed by the stimulus (c.f. figure 14, Chapter 3). Subjects were required to adjust the perceived lateral position of an auditory pointer until they considered it to be in the same position as the stimulus. No time limit was put on the matching process. Each of the 27 different stimuli was presented twenty times in twenty blocks. Each block contained one example of each of the 27 stimuli in a different random order. Rest intervals were permitted after every 180 trials.

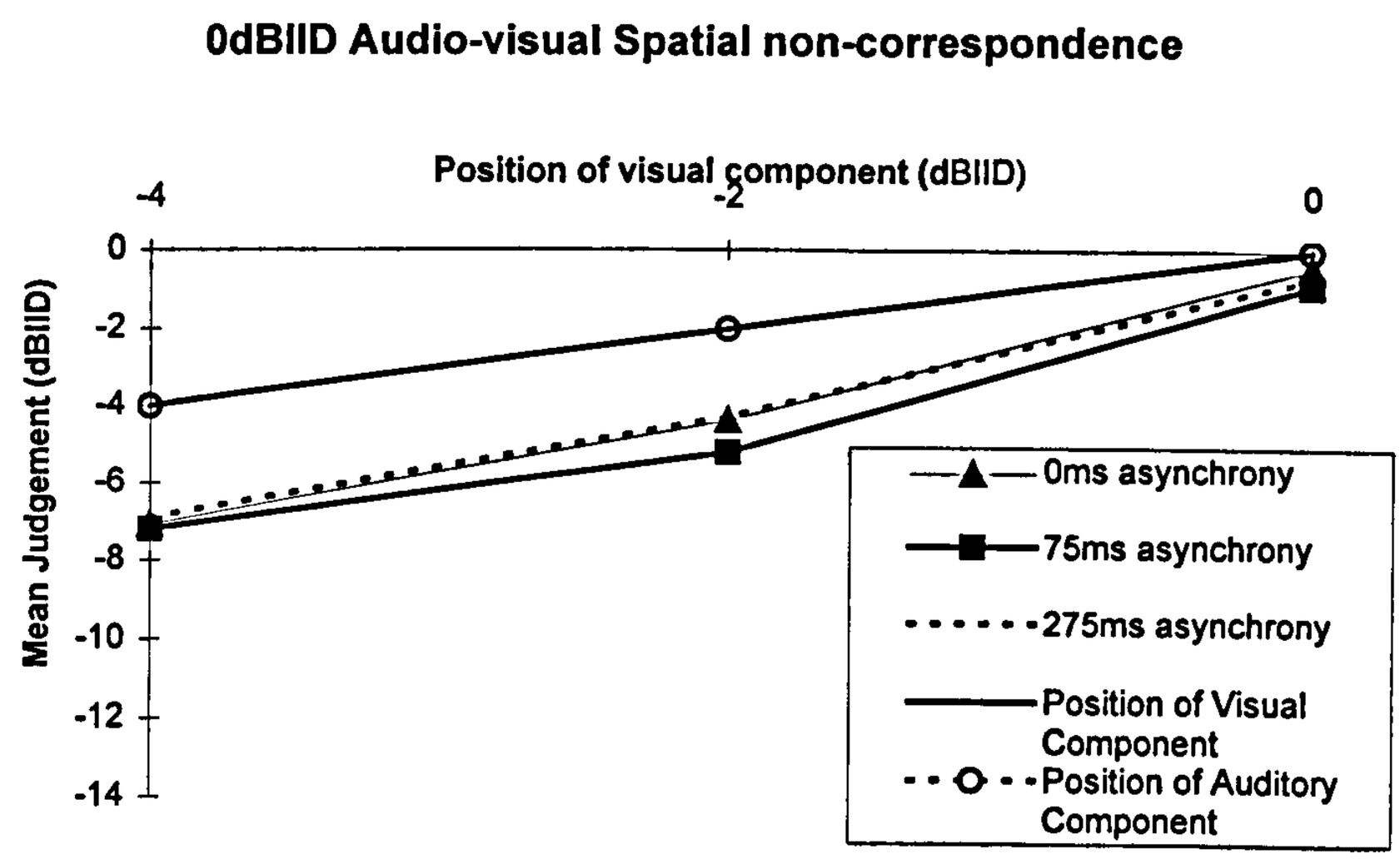
10.8 RESULTS

Mean lateralisation responses as a function of visual presentation position at each of the three levels of audio-visual spatial discrepancy are shown in figure 33. The positions of the auditory and visual components of the audio-visual stimuli are included in each figure. The figures suggest an influence of the auditory component, with mean judgements of position increasing in eccentricity as the position of the auditory component became more eccentric.

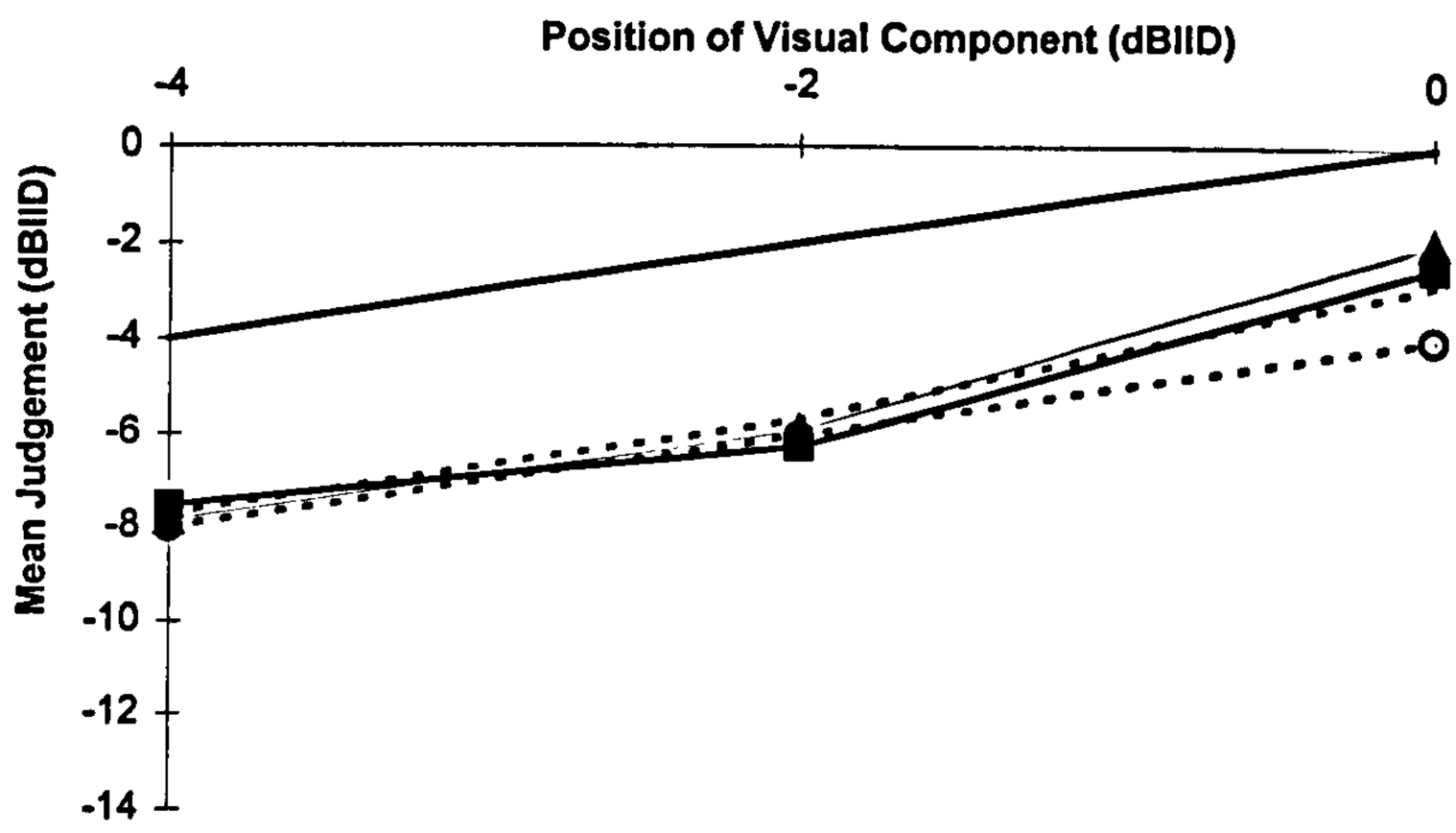
A 3-way analysis of variance with audio-visual asynchrony (3 levels), visual presentation position (3 levels) and audio-visual spatial non-correspondence (3 levels) as factors showed no significant main effect of audio-visual asynchrony [$F(2,10)=0.19$, $p=0.828$]. A significant main effect of visual presentation position was shown [$F(2,10)=55.4$, $p<0.001$] and the influence of spatial difference was consistent but not quite significant [$F(2,10)=4.02$, $p=0.052$].

FIGURE 33 MEAN JUDGEMENTS OF AUDIO-VISUAL STIMULI WITH COMPONENTS VARYING IN ASYNCHRONY AT EACH LEVEL OF AUDIO-VISUAL SPATIAL NON-CORRESPONDENCE

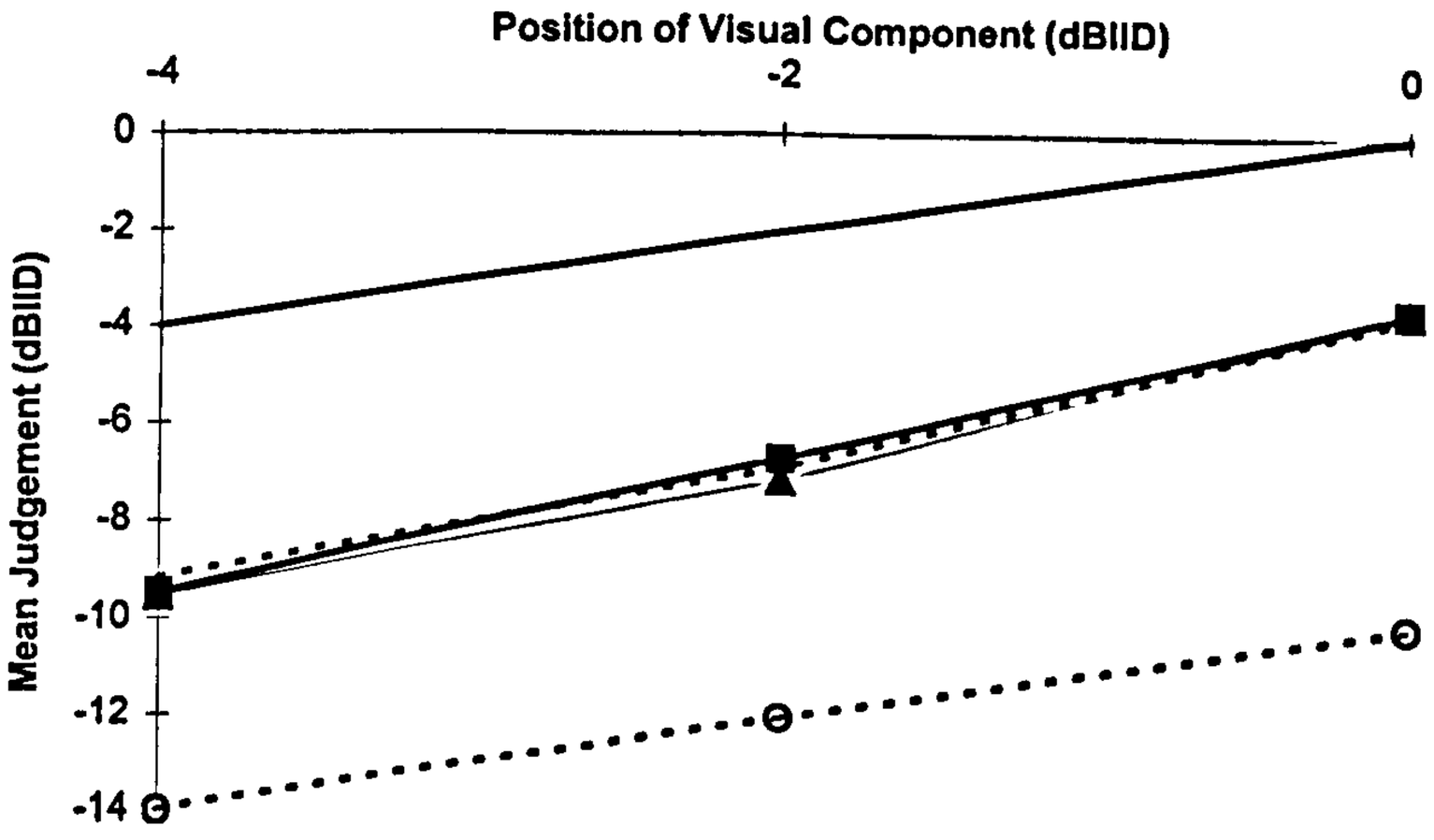
33a



33b



33c



Means of values at each visual position in figures 33 a, b and c are plotted in figure 34. The significance of presentation position is shown by the negative gradient of the functions. The vertical separation of the functions highlights the effect of audio-visual spatial difference.

Mean standard deviations as a function of presentation at each of the three levels of audio-visual spatial mismatch are plotted in figures 35 a, b and c. No significant main effects of audio-visual spatial difference [$F(2,10)=1.35$. $p=0.302$], audio-visual asynchrony [$F(2,10)=0.46$, $p=0.643$] or visual presentation position [$F(2,10)=1.66$, $p=0.239$] were shown by ANOVA. There was a significant interaction between visual presentation position and audio-visual asynchrony [$F(4,20)=3.21$, $p=0.034$] and a three-way interaction between audio-visual spatial difference, audio-visual asynchrony and visual presentation position [$F(8,40)=2.19$, $p=0.049$]. Mean standard deviations in lateralisation judgements collapsed across audio-visual spatial mismatch are shown in figure 36. The significant interaction between visual presentation position and audio-visual asynchrony is a result primarily of the tendency for greater consistency in responses to asynchronous stimuli at the -4 dBIID position.

FIGURE 34. Mean judgements of lateral position as a function of presentation position.

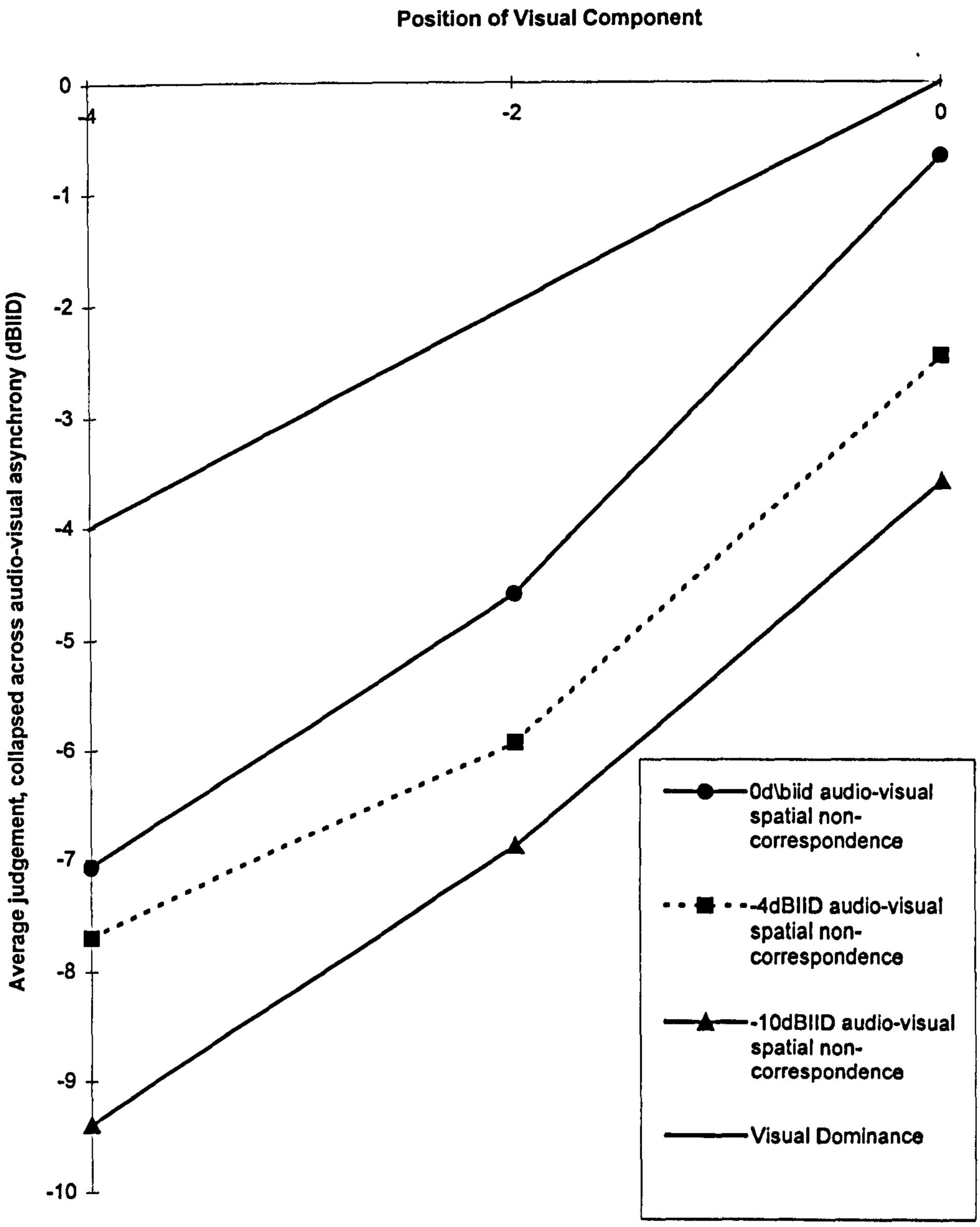
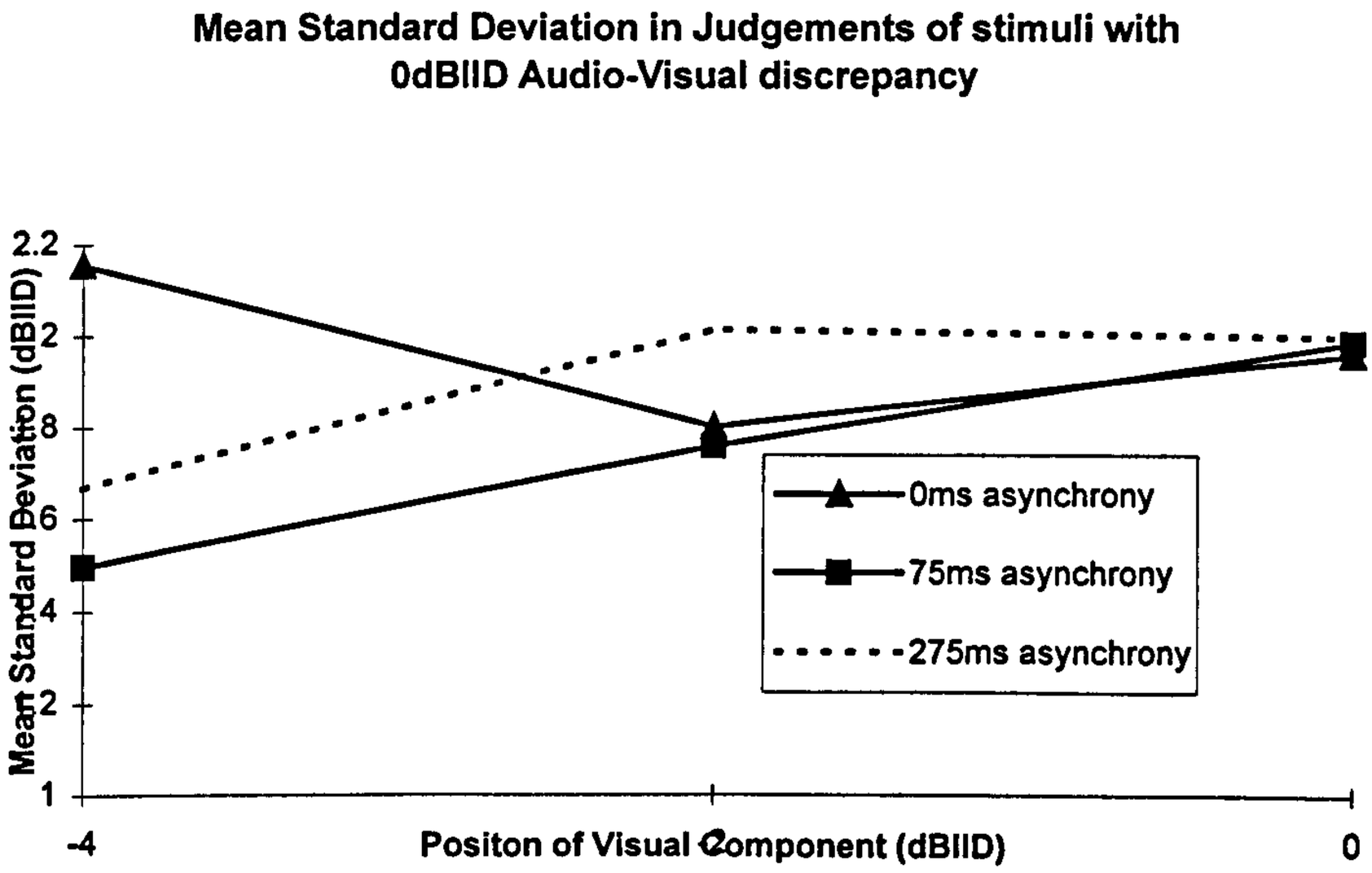
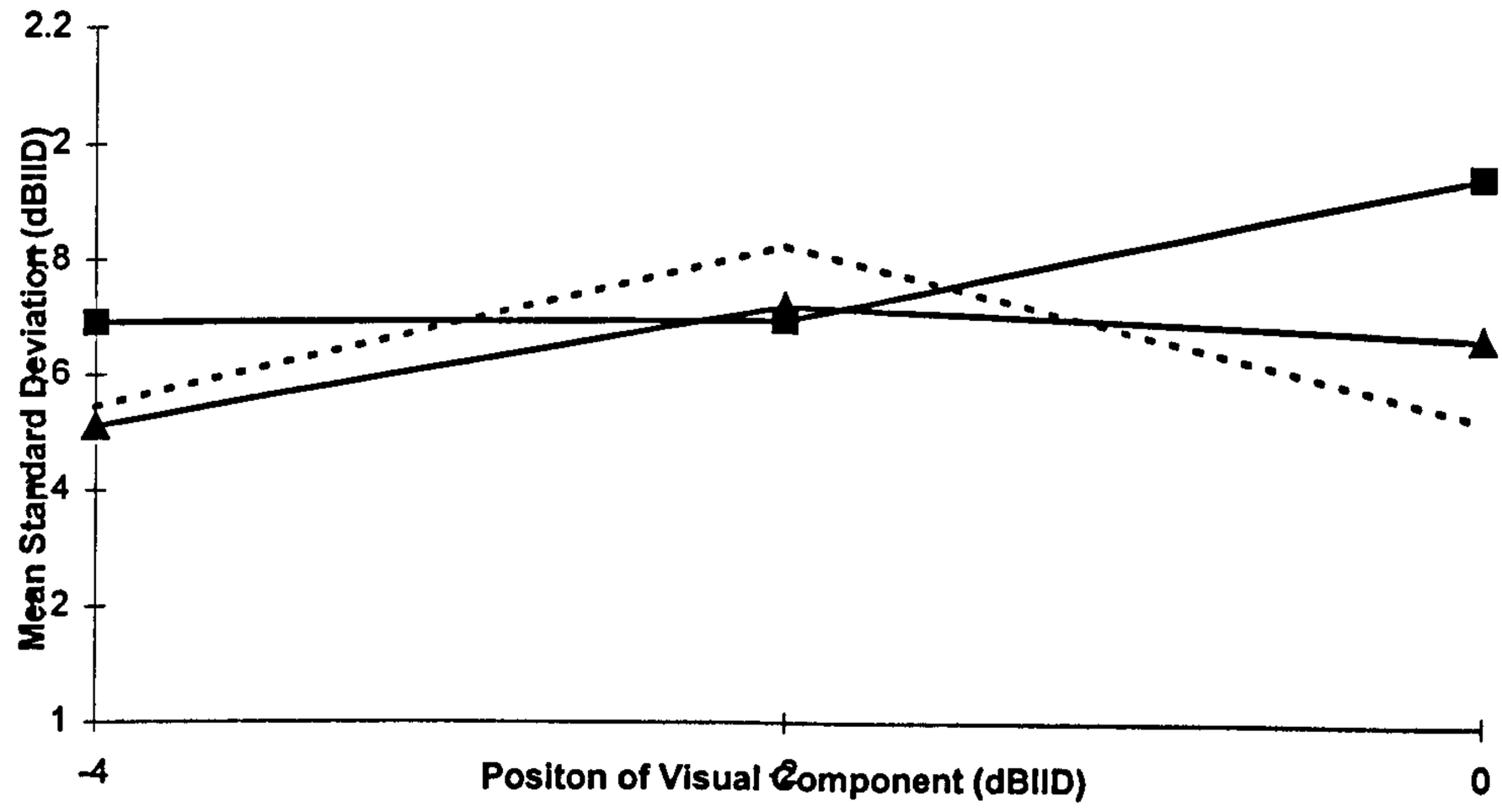


FIGURE 35 MEAN STANDARD DEVIATIONS IN JUDGEMENTS OF AUDIO-VISUAL STIMULI WITH COMPONENTS VARYING IN ASYNCHRONY AT EACH LEVEL OF AUDIO-VISUAL SPATIAL NON-CORRESPONDENCE

35a



35b



35c

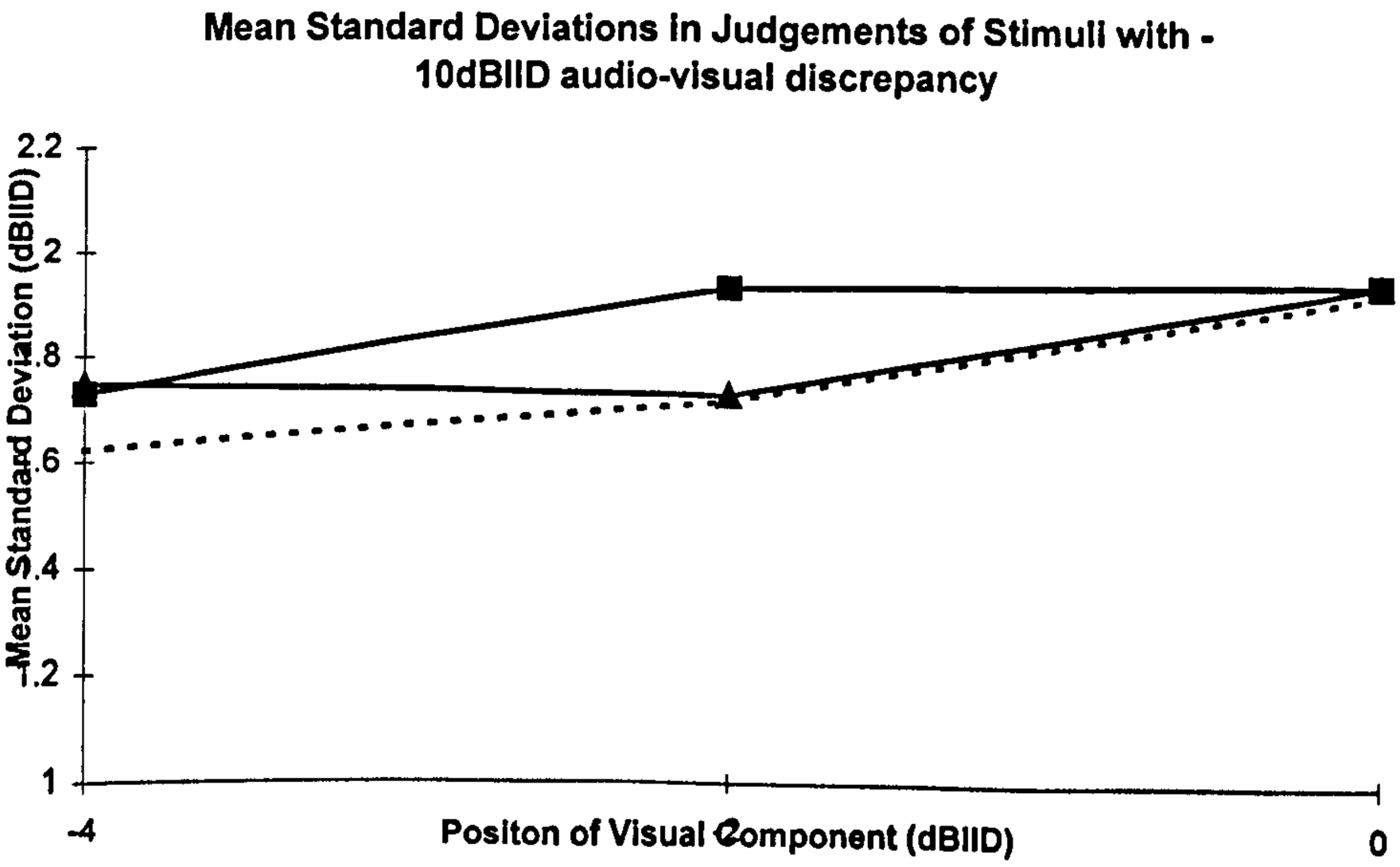
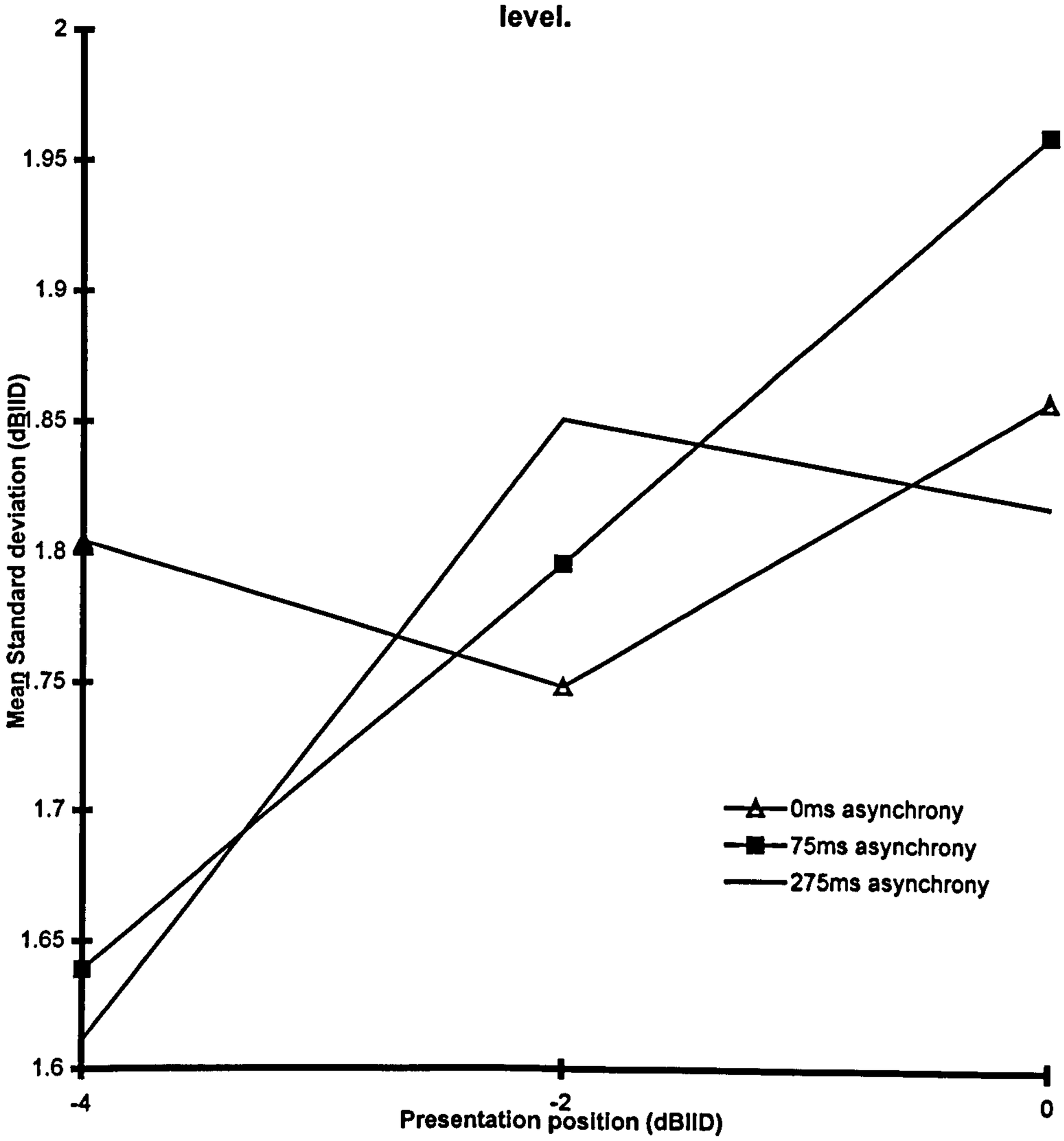


FIGURE 36 Mean standard deviations in lateralisation judgements collapsed across audio-visual spatial mismatch level.



10.9 DISCUSSION

10.9.1 Mean lateralisation judgements (Figure 33 and figure 34)

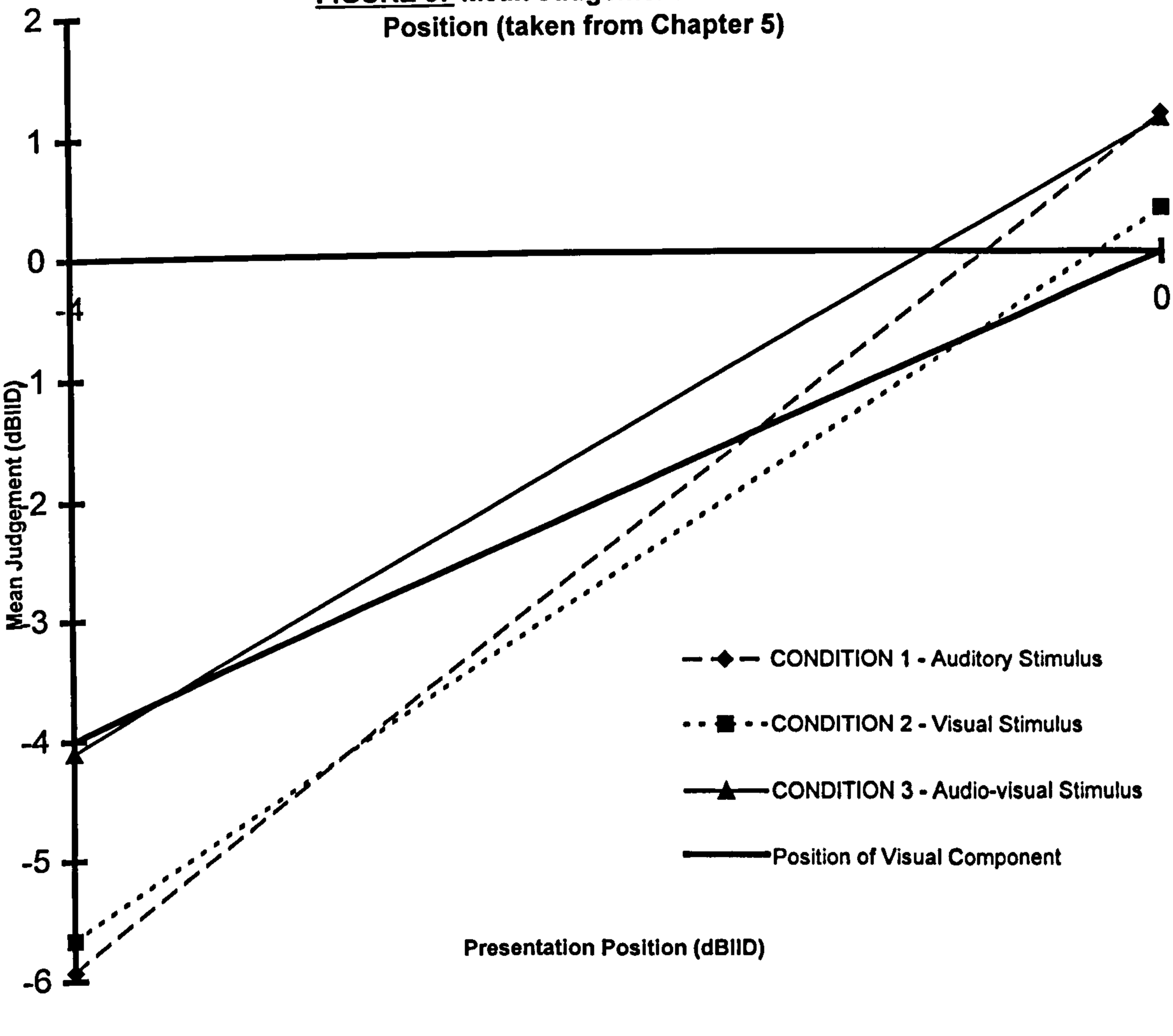
Mean lateralisation judgements (figures 33 a, b and c) indicated that responses were influenced by the the position of the visual component [$F(2,10) = 55.4$, $p<0.001$], shown by the positive gradient of the functions in figure 34. Audio-visual spatial non-correspondence strongly influenced lateralisation judgements [$F(2,10) = 4.02$, $p<0.052$], as indicated by the vertical separation of the individual functions plotted in figure 34. ANOVA indicated that audio-visual asynchrony did not significantly affect mean responses, consistent with the results of the previous experiment in which lateralisations of audio-visual stimuli with temporally non-corresponding components were investigated, although components in the previous experiment did not vary in spatial correspondence.

Previous experiments have all shown mean responses in the position of the visual component, a result suggesting the relative dominance of the visual modality. In this experiment mean responses varied with the position of the visual stimulus component as before, but were biased towards the left of the visual component. Mean position judgements became more eccentric as audio-visual spatial mismatch increased. This tendency for judgements of

position to be influenced by the position of the auditory component of an audio-visual stimulus was not evident in the data from previous experiments.

Figure 33a shows mean responses to stimuli with auditory and visual components in analogous spatial positions. The figure shows that stimuli with spatially and temporally corresponding components fell below the notional line of visual dominance, reflecting a tendency for responses to be more eccentric than the stimulus position. The experiment discussed in chapter 5 compared auditory lateralisation responses to audio-visual stimuli, auditory stimuli, and visual stimuli, in which the components of the audio-visual stimulus corresponded spatially and temporally. Mean responses are plotted in figure 17 , Chapter 5. A section of figure 17, referring in part to stimuli similar to those presented in this experiment is replotted in figure 37 below.

FIGURE 37 Mean Judgements of Stimulus Position (taken from Chapter 5)



The results of the experiment described in chapter 5 showed that mean responses were independent of whether stimuli were presented visually, auditorily or audio-visually, although variance in responses to audio-visual stimuli was found to be lower than variance in responses to uni-modal stimuli. Figure 37 (filled triangles) shows mean position responses in the experiment described in chapter 5 to stimuli with neither spatial nor temporal audio-vidual discrepancies – analagous to the data shown in figure 33a (filled triangles). In chapter 5, mean judgements of audio-visual stimuli were near to the position of the stimulus components. In the present experiment, mean responses fell to the left of the actual position of individual modal components, echoing a similar tendency in responses to the uni-modal auditory and uni-modal visual stimuli at -4 dBIID, in the experiment described in chapter 5. This is broadly consistent with the hypothesis that varying both the spatial correspondence and temporal synchrony of the auditory and visual components of the stimulus reduced the assumption of unity (AOU) due to an increase in stimulus unpredictability (to be discussed later in this section). As a consequence of their weakened AOU, subjects mean responses were more characteristic of their having been presented with a uni-modal visual stimulus, rather than an audio-visual stimulus.

The leftward bias may have been a function of the direction of mismatch. Auditory stimuli were always mismatched to the left of the visual stimulus, and responses may have reflected this. Figure 32 also shown that the leftward

bias increased as visual presentation position becomes more eccentric. Lewald and Ehrnstein (1996), - discussed in chapter 6 - showed that the perceived lateral position of an auditory stimulus was further to the left if gaze was directed to the left, and further to the right if gaze was directed to the right, and it may be that the increase in leftward bias as a function of the eccentricity of the visual component reflect these findings. The more eccentric the position of the visual stimulus, and therefore the gaze of the subject, the further left the perceived position of the auditory component. This analysis of the data suggests that the magnitude of the leftward bias is a function of the position of the auditory component. The role of the position of the auditory component is clearer when the data shown in figure 33a are compared with data shown in figures 33b, 33c and in figure 34. The finding that audio-visual spatial mismatch had a significant influence on mean judgements of lateral position did not correspond with the results of the experiments described in chapter 7 in which responses to audio-visual stimuli with spatially corresponding, and spatially non-corresponding auditory and visual components were compared. The results of the experiment described in chapter 7 showed a dominance of the visual modality with mean responses falling in the position of the visual component irrespective of the position of the auditory component (Figures 21 and 22, Chapter 7) Again, the results of this experiment are consistent with the hypothesis that unpredictability in the stimuli, arising because of the co-variation of two structural variables (audio-visual spatial and temporal correspondence) led to a weakening of the AOU (discussed later in this section) and ultimately to the differences between the data in the present

experiment and the data in previous experiments where either the temporal structural variable or the spatial structural variable was varied, but not both simultaneously.

10.9.2 Mean Accuracy Data (Figures 35a, b and c)

Mean standard deviations as a function of presentation position at each of the three levels of audio-visual spatial non-correspondence are plotted in figures 35a, b and c. Relatively constant levels of variance were shown at all presentation positions and at all levels of audio-visual spatial and temporal mismatch. No main effects of spatial mismatch, temporal mismatch or presentation position were shown by ANOVA. The data were in line with predictions made in an earlier section which suggested that no differences in response variance as a function of audio-visual spatial and temporal mismatch would be found. A significant interaction between position of presentation and audio-visual asynchrony was found [$F(4,20)=3.21$, $p=0.034$]. A significant 3-way interaction between audio-visual asynchrony, audio-visual spatial mismatch, and presentation position was also shown [$F(8,40)=2.19$, $p=0.049$.

In previous experiments, only one structural variable had been varied, the other remaining audio-visually consistent throughout the experiment. If this provided subjects with sufficient stimulus predictability to provide 'historical' evidence that the auditory and visual components of the stimulus referred to

the same perceptual event, a relatively strong AOU would have been maintained. As the audio-visual mismatch in the manipulated structural variable increased variance in response increased, but mean responses remained in the position of the relatively dominant visual modal component. By this reasoning, the increase in variance could be described as a result of a lowered AOU, itself a result of increased audio-visual mismatch. Since one structural component remained audio-visually consistent between each stimulus, the level of stimulus predictability should have been higher than in this experiment where both structural variables were manipulated in each stimulus, leading to lower levels of stimulus predictability overall. The relatively high response variance at all levels and for both types of audio-visual mismatch is consistent with the weakened AOU expected given greater stimulus unpredictability.

The two significant interactions [$F(4,20)=3.21$, $p=0.034$; $F(8,40)=2.19$, $p=0.049$] suggest that the accuracy data shown in figures 35 a, b and c were a result of the simultaneous manipulation of three variables. This is consistent with the suggestion that stimulus unpredictability, increased by the simultaneous manipulation of the structural variables is an important factor in these data.

It is also possible that stimuli with a particular configuration of audio-visual temporal and spatial mismatch may give the perception of a moving stimulus. That is to say, the onset of an auditory component lagging behind the onset of

a visual component by a particular amount, spatially mismatched from the visual component by a particular amount may appear to have moved from the position of the visual component to its present position. The stimuli might be described as exhibiting a cross-modal audio-visual apparent motion. If this was the case, then this may account for the significant interactions since specific levels of audio-visual spatial and temporal mismatch need to be combined in order for motion to be perceived.

It has been suggested above that stimulus unpredictability can also account for some of the unpredicted aspects of the mean lateral position data. The relatively high levels of stimulus predictability in experiments where only one structural variable had been varied provided subjects with evidence on which to base an AOU. In the present experiment, lower levels of stimulus predictability facilitated a weaker AOU. The weaker the AOU, the weaker the relative dominance of one modality over another. It follows that the weaker the AOU the stronger the relative influence of the otherwise less influential component. This is consistent with the theories of modal dominance described earlier.

The Modality Appropriateness Hypothesis (MAH), Modality Precision Hypothesis (MPH) and Directed Attention Hypothesis (DAH) all suggest reasons why one modality may be relatively dominant in any particular multi-modal situation. All are based on the premise that the task can be completed successfully with reference to information provided in either modality. A

reduction in an AOU describes a situation in which the perceiver regards it as less likely that information in the individual modalities derives from the same perceptual event. In such a case, reference to one modality alone might not provide sufficient information for the successful completion of the task in hand. Welch and Warren's 'New View of Intersensory Bias' (1982) indicates that the level of intersensory bias (the relative modal dominance) is dynamic, and directly related to the strength of the unitary assumption. A reduction in the AOU would be a causal factor in a change in the balance of relative modal dominance.

By this reasoning, the unpredicted features of the mean position data in this experiment can be partially ascribed to an increase in the relative dominance of the auditory component as a function of a reduced AOU.

Chapter 11

11.0 CONCLUSIONS

The experiments described in this thesis are an investigation into the relative influences of the auditory and visual components of the stimulus on lateralisation judgements of audio-visual stimuli. The effects of audio-visual temporal correspondence and audio-visual spatial correspondence on the relative dominance of each modality have been investigated, and the interaction between these two variables assessed. The results of each of the experiments are summarised below, followed by a general discussion of the results and their implications. Finally, specific suggestions for further research are made, with references to potential applications of the area of research addressed in this thesis.

11.1 Summary of results

11.1.1 Chapter 2 - Preliminary investigation of audio-visual interaction: Lateral tracking of uni-modal and bi-modal stimuli.

It was the objective of the experiment described in chapter 2 to establish whether lateralisation judgements of audio-visual stimuli were more or less accurate than lateralisation judgements of auditory or visual stimuli. The subjects' task was to estimate how far an auditory, visual or audio-visual stimulus of constant velocity would have travelled over a given period of time.

Mean judgements were independent of whether subjects had been presented with auditory, visual or audio-visual stimuli. Mean standard deviations in judgements (SDs) were significantly smaller for judgements of audio-visual stimuli than they were for judgements of auditory or visual stimuli. It was concluded that the SD "tag" was useful in distinguishing between the relative dominance of the modalities in a multi-modal context (c.f. Warren et al 1982). It became evident that response modality should be considered as a possible factor in lateralisations of auditory, visual and audio-visual stimuli. It was possible that subjects could have completed the task by using only temporal information in the stimulus, and it was concluded that the relative influences of the temporal and spatial factors in lateralisation judgements should be assessed before their interaction in a spatially and temporally demanding task was investigated.

11.1.2 Chapter 4 - Lateralisation of stationary auditory and visual stimuli using auditory and visual pointers.

The general procedure and stimulus characteristics were outlined in chapter 3. It was the objective of the experiment detailed in chapter 4 to investigate the relationship between the lateral position of auditory and visual stimuli and their perceived lateral position using auditory and visual pointers. The results showed that mean judgements of the position of stimuli were independent of stimulus and response modality. A linear relationship between mean judgements of stimulus position and the position of the stimulus was confirmed (c.f. Yost 1981). The results demonstrated a correspondence

between visually and auditorily presented lateral positions, allowing the presentation of audio-visual stimuli with laterally corresponding modal components in later experiments. The results indicated that although mean judgements of lateral position were independent of stimulus and response modality, mean accuracy in judgements - measured in mean standard deviations - differed as a function of response modality as had been suggested in chapter 2. The results also suggested that stimulus modality should be considered as a factor in mean judgement accuracy in cases where stimulus and response modalities were in different modalities. It was concluded that since response modality influenced the level of response accuracy it should be kept constant in the investigation of the relative influence of different stimulus characteristics in lateralisation judgements.

11.1.3 Chapter 5 - Lateralisation of audio-visual stimuli with spatially and temporally-corresponding modal components.

The procedure was identical to that used in the experiment detailed in chapter 4. Comparisons were made of auditorily-made lateralisation judgements of auditory stimuli, visual stimuli, and audio-visual stimuli with spatially corresponding auditory and visual components. Mean judgements of lateral position were independent of stimulus type but mean accuracy in judgements - mean SD - was significantly greater (lower mean SDs) for judgements of audio-visual stimuli than for judgements of auditory stimuli, but not significantly different from mean accuracy in judgements of visual stimuli. It was concluded (c.f. Warren et al 1982) that the SD 'tag' indicated a relative

dominance of the visual modality in lateralisations of audio-visual stimuli with spatially and temporally corresponding auditory and visual components in this context. However, the mean accuracy in judgements of audio-visual stimuli was numerically greater for judgements of audio-visual stimuli than either visual or auditory stimuli, suggesting an influence of both the auditory and visual components in audio-visual lateralisations.

11.1.4 Chapter 6 - The audio-visual spatial relationship.

The objective of measurements described in chapter 6 was to quantify the detectability of audio-visual spatial non-correspondence, enabling the presentation of audio-visual stimuli with auditory and visual components differing spatially at known detectabilities. A 2I-2AFC procedure was employed where subjects indicated the interval with spatially corresponding modal components. Results showed that the predicted inverse relationship between audio-visual spatial non-correspondence and errors in the detection of the non-correspondence broke down when the auditory component was mismatched relative to the visual component towards intra-cranial center (ICC). Possible reasons for the anomaly were suggested. It was concluded that future experiments with spatially non-corresponding stimuli should use stimuli with auditory components mismatched relative to visual components away from ICC, ensuring that detectability of the mismatch was positively related to the level of audio-visual spatial non-correspondence. The results provided evidence that the position of the auditory component relative to the

position of the visual component is relevant in discriminations of this kind, and that the auditory component is not ignored in favour of the visual component.

11.1.5 Chapter 7 - Effect of audio-visual spatial non-correspondence on lateralisation judgements.

The procedure was the same as that used in the experiment described in chapter 5. Results showed that lateralisations of audio-visual stimuli with spatially non-corresponding auditory and visual components indicated the stimulus' position as being in the position of the visual component irrespective of the position of the auditory component (c.f. Jackson 1952, Welch and Warren 1981). Variability in response (mean SD's) increased as a function of audio-visual spatial non-correspondence. The results provided more evidence to suggest that the auditory component was not ignored in favour of the relatively dominant visual component. It was concluded that the position of the auditory component relative to the position of the visual component influenced variability in lateralisation judgements.

11.1.6 Chapter 8 - The audio-visual temporal relationship.

The objective of measurements described in chapter 8 was to quantify the detectability of audio-visual temporal non-correspondence, enabling the presentation of audio-visual stimuli with auditory and visual components

differing temporally at known detectabilities. A 2I-AFC procedure was employed where subjects indicated the interval with temporally non-corresponding modal components. Results showed the predicted positive relationship between audio-visual temporal non-correspondence and the detectability of the non-correspondence at asynchronies greater than 50ms. At asynchronies of 50 ms detectability of the asynchronous alternative was at approximately 18%, considerably less than chance. Possible reasons for the low level of performance were suggested in terms of a visual lag (c.f. Poppel 1985, 1988 and Neimi and Naatanen 1981). The data indicated that the average 75% temporal asynchrony detection level was at 125ms.

11.1.7 Chapter 9 - Effects of audio-visual temporal non-correspondence on lateralisation judgements.

Lateralisation judgements of visual stimuli were compared with lateralisation judgements of auditory and visual stimuli with auditory and visual components varying in asynchrony. Results were consistent with the experiment described in chapter 7 in that mean judgements were in the position of the visual component independent of the level of audio-visual non-correspondence. Mean accuracy in judgements of synchronous audio-visual stimuli was greater than mean accuracy in judgements of audio-visual stimuli with asynchronous components and uni-modal visual stimuli. The role of temporal asynchrony in the formation of the AOU - assumption of unity - was discussed.

11.1.8 Chapter 10 - The audio-visual spatio-temporal relationship.

It was the objective of this experiment to investigate the perceptual interactions of the spatial and temporal relationships between the auditory and visual components of an audio-visual stimulus. Subjects made lateralisation judgements of audio-visual stimuli with components simultaneously varying in temporal and spatial correspondence. The results indicated that mean judgements showed an influence of the position of the auditory component of the stimulus, a result which did not correspond with the experiments described in chapters 7 and 9, in which the spatial or temporal correspondence of the components was varied. The results of this experiment suggested that the simultaneous variation of audio-visual temporal and spatial correspondence affected the relative influence of the auditory component. Mean accuracy in lateralisation judgements was independent of the level of audio-visual spatio-temporal non-correspondence. Stimulus unpredictability was discussed as a factor in the mean lateralisation judgement and mean accuracy data.

11.2 GENERAL DISCUSSION

The experiments described in chapters 2 to 10 constitute an investigation of the lateralisation of audio-visual stimuli, and the relative influences of the auditory and visual components of the audio-visual stimulus. The experiments were motivated by a desire to investigate the relative influence of the individual modalities in different multi-modal contexts. In this thesis the context was altered by varying the type of stimulus (auditory, visual or audio-visual), the type of pointer (auditory or visual) and the auditory and/or temporal correspondence of the auditory and visual components of the audio-visual stimulus. The spatial and temporal correspondence of modal components of a multi-modal stimulus have been described as 'structural factors' influencing the strength of the assumption of unity (AOU) - the subject's assumption that the auditory and visual stimulus components derive from the same multi-modal event. The change in the relative influences of the auditory and visual modalities as a function of systematic variation in the spatial and/or temporal correspondence of the auditory and visual components has been discussed in terms of a potential characteristic of a weakened AOU. Mean lateralisation judgements and the variation in lateralisation judgements were both used as metrics for assessing the relative influences of the individual modal components of the stimulus (c.f. Warren et al 1982).

Mean lateralisation judgements of audio-visual stimuli in all but the experiment discussed in chapter 10 suggested a relative dominance of the

visual component of the stimulus. This is consistent with the previous research (described in an introductory section) into the ventriloquist effect (i.e. Jackson et al 1953; Radeau and Bertelson 1977) which showed a propensity for subjects to indicate the apparent source of an audio-visual stimulus with spatially non-corresponding auditory and visual components as being in the position of the visual component.

Presentations of visual stimuli were preceded by a visual cue in the position of the stimulus (figure 14 - chapter 3). It is possible that the cue may have served to bias lateralisation judgements in favour of the position of the visual component. However, the visual cuing procedure was used in all experiments detailed in chapters 3 to 10. Comparisons of the relative influences of the auditory and visual components as functions of the manipulation of the independent variables are therefore valid, irrespective of any possible visual bias produced by the cue.

In the experiment detailed in chapter 10, comparisons of lateralisation judgements of audio-visual stimuli with spatio-temporally non-corresponding auditory and visual components were made. Mean judgements showed an influence of the position of the auditory component suggesting that simultaneous variation of the temporal and spatial structural variables had a greater influence on the relative dominance of the modalities than the individual variation of the spatial or temporal structural factors. The results of the experiment described in chapter 10 suggest that the simultaneous

variation of the spatial and temporal correspondence of the modal components of the stimulus weakened the AOU sufficiently so that the increased relative influence of the auditory component was revealed in mean lateralisation judgements. This implicates the unpredictability of the stimulus, which varies in these experiments as a function of spatial and temporal variation, as a possible influence on the relative dominance of the auditory and visual modalities.

Whereas mean judgements of lateral position in all but the experiment described in chapter 10 showed no influence of the position of the auditory component, the mean accuracy of lateralisation judgements - measured in mean standard deviations - of audio-visual stimuli with spatio-temporally corresponding auditory and visual components was greater than the mean accuracy in lateralisation judgements of uni-modal auditory or visual stimuli. It can be concluded that the auditory and visual components of the stimulus were combined to provide a percept which afforded more accurate lateralisation judgements than would have been possible if only auditory or visual stimuli were available. The experiments described in chapters 7 and 9 suggested that the mean accuracy in judgements of spatially or temporally non-corresponding stimuli decreased as a function of the level of audio-visual non-correspondence. Reduced mean accuracy in lateralisation judgements is a characteristic of judgements of uni-modal rather than audio-visual stimuli (c.f. chapter 5). The increase in mean standard deviations with the level of audio-visual non-correspondence in chapters 7 and 9 may have been a function of a

reduced AOU. It is possible that judgements became more characteristic of uni-modal judgements as the audio-visual non-correspondence increased, and the AOU weakened.

Three theories of multi-modal integration (as described by Welch and Warren, 1982) were described in the introductory section of this thesis. The modality appropriateness hypothesis (MAH), the modality precision hypothesis (MPH), and the directed attention hypothesis (DAH) variously predict that the relatively dominant modality in a multi-modal context would be the modality which was more 'appropriate' (MAH) or 'precise' (MPH), or the modality to which more attention was directed (DAH). Welch and Warren also proposed 'a new view of multi-sensory integration' which accommodated the three existing theories into a model within which the AOU has a central role (chapter 1). Welch and Warren's model suggests that in situations where there is a bias of one modality over another, subjects must have formed a sufficiently strong AOU about the auditory and visual components of the stimulus, and the information in the stimulus streams must be discrepant, although later research (i.e. Warren et al. 1982, and the experiments reported in this thesis) indicates that intermodal bias can be measured in the perception of non-discrepant multi-modal events. The model predicts that if the intermodal discrepancy is too great, or the AOU is too weak, the bias will still exist, but the perceiver will detect a discrepancy in the individual modalities. They suggest that it is possible for each modality to bias the other, and the two

bias effects can be as expressed as percentage biases of one modality of the other.

Welch and Warren describe the following example (c.f. Hay et al. 1965). Let V(P) = visual bias of proprioception. Let P(V) = proprioceptive bias of vision. Subjects provide data in four tasks. In the two control tasks subjects point with their right index finger, beneath a table top, to their unseen left index finger and the error is recorded as Pc, and to a visual target while wearing a displacing prism, the error recorded as Vc. The two control measures provide a baseline level of accuracy in pointing at either a visually displaced object or an unseen proprioceptive target. In the two experimental sessions, subjects viewed their target finger briefly through the displacing prism. They were instructed to point to where they saw the finger, (with the error recorded as Ve) or where they felt it to be (Pe).

$$V(P) = \frac{Pc - Pe}{Pc - Vc} \quad X \quad 100$$

$$P(V) = \frac{Vc - Ve}{Vc - Pc} \quad X \quad 100$$

In the experimental sessions both visual and proprioceptive information was available to the subject. The difference between performance on the proprioceptive experimental task (Pe) and the proprioceptive control task (Pc) provides a measure of the influence of the visual modality. When taken as a ratio of the difference in performance on the two control tasks, a value

representing the percentage bias of the visual modality over the proprioceptive modality is obtained.

Welch and Warren indicate that with a strong AOU in a task like the one described, the sum of the two perceptual biases typically approximates to 100%, and the observer will detect no discrepancy - intermodal non-correspondence - in the individual modalities (c.f. Warren and Pick 1970). With a weaker AOU, the sum of the two bias effects is typically less than 100%, and there is the possibility that the residual intermodal non-correspondence will be detected.

The model allows the possibility that the level of intersensory bias of one modality of another is a continuum rather than a categorical measure. In this thesis, the reduction in the influence of the visual modality - the increase in the relative influence of the auditory modality - as a function of an increase on audio-visual spatial or temporal non correspondence (Chapters 6 and 8) has been described in terms of the mean accuracy in lateralisation judgements rather than a change in the percentage bias of the visual modality over the auditory modality, but the end result is essentially the same. These results, and Welch and Warren's model both suggest a continuum between the absolute dominance of one modality in a bi-modal scene and the absolute dominance of the other modality. Similarly both suggest that movement along the continuum is facilitated by a change in the strength of the AOU, possibly as a function, or a symptom of increased intermodal non-correspondence. In

these respects, the results of the experiments described in chapters 6 and 8 in this thesis provide support for Welch and Warren's 'new view of intersensory bias'.

The nature of the relationship between the AOU and the structural correspondence of the individual modal elements of a multi-modal stimulus is difficult to define. On one hand, the strength of the AOU is partly a function of the structural correspondence of the individual modal streams. On the other hand, the level of structural non-correspondence which can be tolerated before it is detected is partly a function of the strength of the AOU. In this sense, and in this context the AOU and the level of cross-modal correspondence are synonomous. A differentiation between the AOU and the level of intermodal structural correspondence is possible when other factors are considered. The AOU, for instance, is influenced by cognitive variables, including past experience with the stimulus, factors which do not influence the structural correspondence of the modal components.

The results of the experiment described in chapter 10 suggested no clear intersensory bias of either modality, although the results did suggest an increase in the relative influence of the auditory component. Similarly, the increase in the influence of the auditory component was shown with spatio-temporally corresponding, and therefore perceptually corresponding stimuli, a result discussed in terms of stimulus unpredictability. This is an example of an influence on intersensory bias of 'specific historical factors' (Welch &

Warren's term) resulting from experience with, and knowledge of the stimulus ensemble.

The role of experience with the type of stimulus used should be addressed in the context of this thesis. The same group of subjects provided data for all experiments described in chapters 3 to 10. It is likely that during this exposure to the stimuli, subjects became sensitive to auditory and visual temporal and spatial discrepancies which they would otherwise not have perceived, or they may have began to listen to and watch the stimulus more analytically. It is possible that their results, especially on the later experiments in the series, may have reflected this enhanced sensitivity, or more analytical behaviour. In the experiment described in chapter 9, results were characteristic of subjects having identified a 125ms auditory and visual component asynchrony more accurately than earlier difference limen measurements suggested. Also, the sensitivity of mean lateralisation judgements to the position of the auditory component in chapter 10 could have been a function of increased sensitivity to the spatial difference in the auditory and visual components of the audio-visual stimulus, itself a consequence of subjects' experience with the stimulus. However, when subjects were debriefed after the experiment described in chapter 10 they were surprised to discover that the majority of stimuli had spatially and temporally non-corresponding auditory and visual components. They indicated that they were occasionally aware that there was something about the stimuli that they could not articulate, and that large asynchronies or spatial differences were obvious, but none of the subjects was aware that most

stimuli had spatio-temporally non-corresponding components. In future experiments of this kind the role of specific experience with the stimuli could be controlled and investigated by changing the subjects used in each experiment, or by systematically varying each subject's experience with the stimuli.

The effects of audio-visual spatial and/or temporal mismatch on lateralisation judgements of audio-visual stimuli could benefit from further investigation. If the relative influences of the modalities are continuously rather than categorically differentiated, as has been suggested by the results of these experiments, then methods of systematically varying the position of a stimulus on the continuum would allow some more detailed modelling of intersensory interaction. In the experiments described in this thesis, the effects on lateralisation judgements of varying the temporal and/or spatial correspondence on the auditory and visual components were investigated. Future experiments might investigate other variables which may influence audio-visual interaction. These might include visual elevation / auditory pitch height, visual depth and/or size / auditory intensity and any combination of suitable audio-visual pairings, providing data for a detailed model of audio-visual interaction.

The results of the audio-visual spatial and temporal correspondence difference limen measurements (chapters 6 and 8) indicated two related areas which would benefit from further investigation. It may be the objective of an

application, for instance in telecommunications or in aviation based audio-visual displays, to provide subjects with optimal conditions for lateralisation, including mean lateralisation performance and mean accuracy of lateralisation judgements. These experiments have shown that optimal stimuli for tasks such as these are audio-visual stimuli with spatially and temporally corresponding auditory and visual components. However, audio-visual spatial difference limen measurements (chapter 6) suggested that the auditory and visual components of audio-visual stimuli with the auditory component mismatched relative to the visual component by approximately 1dBIID towards ICC were perceived as spatially corresponding more often than similar stimuli with spatially corresponding components. Reasons for this anomaly have been discussed in terms of the influence of past experience and an orienting reflex. The influence of gaze direction on auditory lateralisation on audio-visual stimuli has also been discussed in the context of audio-visual spatial correspondence. Lewald and Ehrnstein (1996) showed that the lateral position of an auditory stimulus is shifted in direction of gaze. The results of the experiment described in chapter 6 suggest that the perception of audio-visual spatial correspondence is enhanced by spatially separating the auditory and visual components as described. The results of Lewald and Ehrnstein (1996) suggest that the degree with which the perceived lateral position of the auditory component is shifted is a function of the eccentricity of the listeners gaze - the more eccentric the gaze direction, the greater the shift in the perceived lateral position of the auditory stimulus. If it is the intention to generate audio-visual stimuli with spatially corresponding auditory and visual

components, and stimuli are to be presented off-center, then the influence of gaze direction on lateralisation as a function of gaze eccentricity should be taken into consideration. Further research is needed into how robust the results shown in chapter 6 are, and if the result occurs or varies at different eccentricities (c.f. Lewald and Ehrnstein 1996). Similarly, audio-visual temporal difference limen measurements (chapter 8) highlighted a visual perceptual lag of approximately 50ms. If it is the intention that auditory and visual components of a visual stimulus are perceived as synchronous, then the visual lag should be taken into consideration. More detailed calculations of the length of the perceptual lag using a similar 2I-2AFC procedure to that used in chapter 8 would provide data enabling perceptually synchronous audio-visual presentations, and advance existing knowledge of audio-visual interactions of this kind. Finally, experimentation into whether the results outlined in this thesis can be extended to judgements of audio-visual stimuli with components varying in elevation as well as azimuth (localisation rather than lateralisation measurements) provides further scope for the investigation of audio-visual interaction.

A greater understanding of audio-visual interaction would benefit the design of systems where auditory and visual devices are used in a confined space, often simultaneously. Warning systems, for instance have traditionally used loud sounds combined with flashing or bright lights to elicit the required response. Unfortunately, as Patterson (1990) reports, the effect is often "exactly what was NOT intended". He reports an incident in which a pilot, when confronted

by numerous auditory and visual warning systems at once, found that he had to cancel the alarms before he was able to concentrate on addressing the problem. Patterson goes on to point out that the emphasis should be on warning the listener rather than startling them. Since multi-modal perceptual information available to a user, for example a driver, is being increased all the time, a better understanding of how visual and auditory information interacts in the perceptual process is required. More recently the specific experience of the user of the alarm system has been considered as a variable in the design of the sound (Edworthy and Stanton 1995), and information-carrying characteristics of an alarm signal in addition to its perceived urgency have been looked into (Hellier and Edworthy 1989). If it is the intention to design a warning device so that the signal is informative about an event, and spatial localisation is important in the context of the alarm, then the results of the experiments reported in this thesis suggest that designers should consider an audio-visual signal with temporally and spatially corresponding auditory and visual components, taking the results of the audio-visual spatial and temporal correspondence DL's into consideration, to ensure optimal localisation accuracy.

In conclusion, the experiments in this thesis have confirmed a relative influence of the visual modality in lateralisation judgements of audio-visual stimuli. However, the results of the experiments demonstrate that neither the position of the auditory component nor its temporal relationship with the visual component is ignored in lateralisation judgements of the audio-visual

stimulus. The results showed that the less the auditory and visual components of the stimulus corresponded structurally (spatially or temporally), the greater the relative influence of the auditory modality. The experiments demonstrate that comparing the mean accuracy of judgements (mean SD's) is a valid method of investigating the relative influence of the corresponding and non-corresponding auditory and visual modalities in an audio-visual context (c.f. Warren et al 1982).

APPENDIX I

EXPERIMENTAL INSTRUCTIONS

This experiment is concerned with your judgements of spatial position. You will be required to respond to a stimulus using an ***Acoustic Pointer****, a tone whose position inside your head can be moved with the track-ball.*

* *At the beginning of each trial, an auditory* ***Stimulus*** *will be presented, a tone over headphones.*

** *At the beginning of each trial, a visual* ***Stimulus*** *will be presented, a spot on a diagram of the back of your head.*

*** *At the beginning of each trial, an audio-visual* ***Stimulus*** *will be presented, a tone via headphones, and a spot on a diagram of the back of your head.*

It is your task to move the ***Pointer*** *to the position of the* ***Stimulus****, and hit the key marked "NEXT" when you are happy with your match. The next trial will follow after a short delay.*

Speed is NOT important, take as long as you need to make the match.

(*/**/*** Depending on the stimuli to be presented)

APPENDIX II

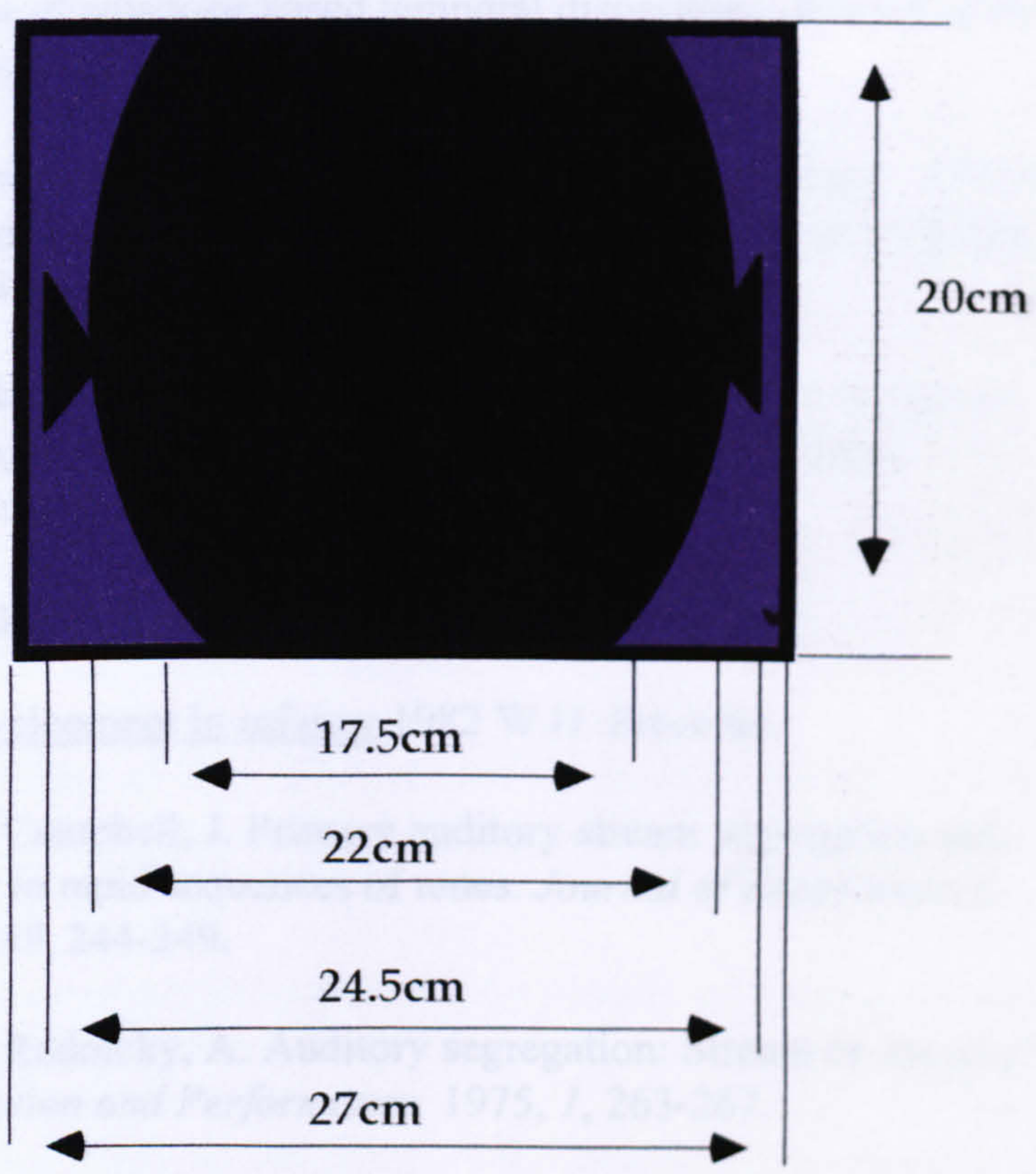


Diagram of Head Silhouette

REFERENCES

Anstis, S. & Saida, M. Adaptation to auditory streaming of frequency modulated tones. *Journal of Experimental Psychology: Human Perception and Performance,* 1985, *11,* 257-271.

Auerbach, L., & Sperling , P.A. A common auditory-visual space: Evidence for its reality. *Perception and Psychophysics,* 1974, *16,* 129-135.

Bernstein, L.R, & Trahiotis C. Lateralisation of low-frequency, complex waveforms: The use of envelope based temporal disparities. *Journal of the Acoustical Society of America,* 1984, *77(5),* 1868-1880.

Bloch, H. Status and function of early sensory-motor coordination . 1990 In sensory-motor organisations and development in infancy and early childhood, Bloch & Bertenthal (Eds.).

Bloch, H. Intermodal participation in the formation of action in the infant. 1994 In. The development of intersensory perception. Comparative perspectives. Lewkowicz D. J., & Lickliter R. (Eds.).

Bower, T. G. R. Human development. 1979 W. H. Freeman.

Bower, T.G.R. Development in infancy 1982 W.H. Freeman.

Bregman, A. S., & Campbell, J. Primary auditory stream segregation and perception of order in rapid sequences of tones. *Journal of Experimental Psychology,* 1977, *89,* 244-249.

Bregman, A. S., & Rudnicky, A. Auditory segregation: Stream or streams? *JEP Human Perception and Performance,* 1975, *1,* 263-267.

Bregman, A. S., Auditory Scene Analysis. The perceptual organisation of sound. 1990, MIT press.

Bregman, A. S., & Pinker, S. Auditory streaming and the building of timbre. *Canadian Journal of Psychology,* 1978, *32,* 19-31.

Bruce, V., & Green, P. Visual Perception: Physiology, Psychology and Ecology. 1987, LEA.

Cherry, E. C. Some experiments on the recognition of speech with one and with two ears. *Journal of the Acoustical Society of America,* 1953, *25,* 975-979

Cutting, J. E., & Rosner, B. S. Categories and boundaries in speech and music. *Perception and Psychophysics,* 1974, *16,* 574-570.

Dannenbring, G. L., & Bregman, A. S. Streaming vs. fusion of sinusoidal components of complex waves. *Perception and Psychophysics,* 1978, *24,* 369-376.

Darwin, C.J. Perceiving vowels in the presence of another sound: Constraints on formant perception. *Journal of the Acoustical Society of America,* 1984, *76,* 1636-1647.

Darwin, C. J., & Ciocca, V. Grouping in pitch perception: Effects of onset asynchrony and ear of presentation of a mistuned component. *Journal of the Acoustical Society of America,* 1992, *91(6),* 3381-3390.

Deutsch, D. Binaural integration of melodic patterns. *Perception and Psychophysics*. 1979, *25,* 399-405.

Dixon, N.F, & Spitz, L. The detection of audiovisual desynchrony. *Perception,* 1980, *9,* 719-721.

Dodd, B. Lip reading in infants: Attention to speech presented in- and out-of-synchrony. *Cognitive Psychology,* 1979, *11,* 478-484.

Dodd, B. Effects of social and vocal stimulation on infant babbling. *Developmental Psychology.* 1972, *7,* 80-83.

Dodd, B. The acquisition of lip-reading skills by normally hearing children. In Hearing by Eye: The Psychology of Lip-Reading. Dodd & Campbell eds. 1987 LEA.

Driver J. Enhancement of selective listening by illusory mislocation of speech sounds due to lip-reading. *Nature.* 1996. *381.* 66-68.

Duhamel, J.-R., Colby, C.L., & Goldberg, M.E. Congruent visual and somatosensory response properties of neurons in the ventral intraparietal area (VIP) in the alert monkey. *Society for Neurosciences Abstracts,* 1989, *15,* 162.

Edworthy J., Stanton N. A user-centred approach to the design and evaluation of auditory warning signals: 1. Methodology. *Ergonomics,* 1995, *38(11),* 2262-2280.

Elliot, R. Simple visual and simple auditory reaction time: a comparison. *Psychonomic Sciencel,* 1968, *10,* 335-336.

Ellis, R.R., & Lederman, S. J. The role of haptic versus visual volume cues in the size-weight illusion. *Perception and Psychophysics,* 1993, *3,* 315-324.

Fishkin, S. M., Pishkin, V., & Stahl, M.L. Factors involved in visual capture. *Perceptual and Motor Skills,* 1975, *40,* 427-434.

Gibson , J. J. The ecological approach to visual perception. 1979 Boston: Houghton-Mifflin.

Gopher D. Eye movement patterns in selective listening tasks of focussed attention. *Perception and Psychophysics,* 1973, *14,* 259-264.

Grantham D.W. Interaural intensity discrimination: insensitivity at 1000Hz. *Journal of the Acoustical Society of America,* 1984, *75(4),* 1191-1194.

Guay M. Short-term retention of temporal auditory information. *Perceptual and Motor Skills,* 1982, *19-26.*

Handel, S., Weaver, M. S., & Lawson, G. L. Effect of rhythmic grouping on stream segregation. *JEP Human Perception and Performance, 1*983, *9,* 637-651.

Hay, J.C. Pick, H. L., Jr., & Ikeda, K. Visual capture produced by prism spectacles. *Psychonomic Science,* 1965, *2,* 215-216.

Held, R. Shifts in Binaural localization after prolonged exposures to atypical combinations of stimuli. *American Journal of Psychology.* 1955. *68.* 526-548.

Hellier E. & Edworthy, J. Quantifying the perceived urgency of auditory warnings, *Canadian Acoustics,* 1989, *17(4),* 3-11.

House A.S., & Stevens K. N. Auditory testind of a simplified description of vowel articulation. *Journal of the Acoustical Society of America.* 1955, *27(5),*882-887.

Jackson, C. V. Visual factors in auditory localisation. *Quarterly Journal of Experimental Psychology,* 1953, 52-66.

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

Kalil, R. & Freedman, S. J. Compensation for auditory rearangement in the abscence of observer movements. *Perceptual and Motor Skills,*1967, *24,*475-478.

King, A. J, & Carlile, S. Changes induced in the representation of Auditory space in the superior colliculus by rearing ferrets with binocular cyclid suture. *Experimental Brain Research,* 1993, *94,* 444-455.

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

King, A.J, & Palmer, A.R. Integration of visual and auditory information in bi-modal neurones in the guinea-pig superior colliculus. *Experimental Brain Research.* 1985, *60,* 492-500.

Knudsen, E.I., Early auditory experience aligns the auditory map of space in the optic tectum of the barn owl. *Science,* 1983, *222*, 939-942.

Knudsen E.I, & Knudsen, P.F. Vision calibrates sound localisation in developing barn owls. *Journal of Neuroscience*, 1989, *9,* 3306-3313.

Knudsen, E.I, & Konishi, M. A neural map of auditory space in the owl. *Science*, 1978, *200*, 795-797.

Koffka, K. Principles of Gestalt Psychology. 1935, New York: Harcourt Brace.

Kubovy, M., Cutting, J. E., & McGuire, R. M. Hearing with the third ear: Dichotic perception of a melody without monaural familiarity cues. *Science N.Y.,* 1974, *186,* 272-274.

Kuhl, P.K. & Meltzoff, A.N. The bimodal perception of speech in infancy. *Science.* 1982, *218,* 1138-1141.

Langsdorf, P, Izard, C., Rayais, M & Hembree, E. Interest expression, visual fixation and heart rate changes in 2 to 8 months old infants. *Development Psychology.* 1983, *3*, 375-386.

Lewald, J. & Ehrnstein, W.H. The effect of eye position on auditory lateralisation. *Experimental Brain Research.* 1996, *108*, 473-485.

Lewkowicz D.J. Perception of auditory-temporal synchrony in Human Infants, *JEP Human Perception and Performance*, 1996, *22(5)*, 1094-1106.

Lewkowicz D. J., & Lickliter R. (Eds.) The development of intersensory perception. Comparative perspectives. 1994 LEA.

Massaro, D. W. Speech perception by ear and eye. In Hearing by Eye: The Psychology of Lip-Reading Dodd & Campbell eds. 1987 LEA.

MacLeod, A. & Summerfield, Q. A Procedure for measuring auditory and audiovisual speech reception thresholds for sentences in noise: rationale, evaluation and recomendations for use. *British Journal of Audiology*,1990, *24,* 29-43.

McDonnell, P.M. & Duffett, J. Vision and Touch: A reconsideration of the conflict between the two senses. *Canadian Hournal of Psychology,* 1972, *26,* 171-180.

McGrath, M. & Summerfield, Q. Intermodal timing relations and audiovisual speech recognition by normal hearing adults. *Journal of the Acoustical Society of America,* 1985, *2,* 678-685.

McGurk, H. & MacDonald, J. W. Hearing lips and seeing voices. *Nature, London,* 1976, *264,* 126-130.

McNally, K. A., & Handel, S. Effect of element composition on streaming and the ordering of repeating sequences. *JEP Human Perception and Performance,* 1977, *3,* 451-460.

Meltzoff, A.N., & Borton R.W. Intermodal matching by human neonates. *Nature,* 1979, *282,* 403-404.

Meltzoff, A. N, & Kuhl, P. K. Faces and speech: Intermodal processing of Biologically relevant signals in infants and adults. 1994 In. The development of intersensory perception. Comparative perspectives. Lewkowicz D. J., & Lickliter R. (Eds.)

Meredith, M.A., & Stein B.E., Interactions among converging sensory inputs in the superior colliculus. *Science.* 1983, *221,* 389-391.

Meredith, M.A., & Stein B.E. Spatial factors determine the activity of multisensory neurons in cat superior colliculus. *Brasin Research.* 1986, *365,* 350-354.

Metelli, F. The perception of transparency. *Scientific American,* 1974, *230(4),* 90-98

Moore, B. C. J. An introduction to the psychology of hearing, 1989 3rd.edition Academic Press.

Morrongiello, B. A. Effects of colocation on auditory-visual interactions and cross-modal perception in infants. 1994 In. The development of intersensory perception. Comparative perspectives. Lewkowicz D. J., & Lickliter (Eds.)

Niemi, P. & Näätänen R. Foreperiod and simple reaction time. *Psychological Bulletin,* 1981, *89(1),* 133-162.

O'Connor N. & Hermelin B. Seeing and hearing in space and time. *Perception and Psychophysics,* 1972, *11(1A),* 46-48.

O'Leary, A., & Rhodes, G. Cross-modal effects on visual and auditory object perception. *Perception and Psychophysics,* 1984, *6,* 565-569.

Patterson R.D. Auditory warning sounds in the work environment. *Philosophical Transactions of the Royal Society of London B.,* 1990, *327,* 485-492.

Piaget, J. The construction of reality in the child. 1954 N.Y.: Basic Books.

Piaget J. The Origins of intelligence in children. 1952 New York: International Universities Press.

Pick, H. L., Jr., Warren, D. H., & Hay, J. C. Sensory conflict in judgements of spatial direction. *Perception and Psychophysics,* 1969, *6*, 203-205.

Plomp, R., & Mimpen, A. M. Improving the reliability of testing the speech-reception thresholds for sentences. *Audiology,* 1979, *18*, 43-52.

Pöppel E (1988) Mindworks: Time and Conscious Experience. New York: Harcourt, Brace & Jovanovich.

Posner, M.I., Nissen, M.J., & Klein, R.M. Visual Dominance: An information processing account of its origins and significance, *Psychological Review*, 1976, *83*, 157-171.

Power, R. P., & Graham, A. Dominance of touch by vision: Generalization of the hypothesis to a tactually experienced population. *Perception,* 1976, *5*, 161-166.

Radeau, M. & Bertelson, P. Adaptation to auditory-visual discordance and ventriloquism in semirealistic situations. *Perception and Psychophysics,* 1977, *22*, 137-146.

Rauschecker, J. P., & Kniepert, U. Auditory localisation behaviour in visually deprived cats. *European Journal of Neuroscience,* 1993, *6*, 149-160.

Rasch, R. A. The perception of simultaneous notes such as in polyphonic music. *Acoustica,* 1979, *40,* 21-33.

Reisberg, D. Looking where you listen: Visual cues and auditory attention *Acta Psychologica,* 1978, *42*, 331-341.

Reisberg, D, Scheiber, R., & Potemkin, L. Eye position and the control of auditory attention. *JEP Human Perception and Performance,* 1981, *7(2)*, 318-323.

Regan, M. P., He, P, & Regan D. An audiovisual convergence area in the human brain *Experimental Brain Research,* 1995, *106,* 485-487.

Roig M. & Cicero F. Hemispherity, style, sex and performance on a line-bisection task: An exploratory study. *Pereptual and Motor Skills.* 1994, *78,* 115-120.

Rutschmann J., & Link R. Perception of temporal order of stimuli differing in sense mode and simple reaction time. *Perceptual and Motor Skills,* 1964, *18,* 345-352.

Saldaña, H., M., & Rosenblum, L. D. Visual influences on auditory pluck and Bow judgements. *Perception and Psychophysics,* 1993, *3,* 406-416.

Schiano J.L., Trahiotis C. & Bernstein L.R. Lateralisation if low-frequency tones and narrow bands of noise, *Journal of the Acoustical Society of America,* 1986, *79(5),* 1563-1570.

Schiff,W. & Detwiler, M.L. Information used in judging impending collisions. *Perception*, 1979, *8,* 647-658.

Scheffers, M. T. M. Sifting Vowels: Auditory pitch analysis and sound segregation. *1983 Unpublished Doctoral Dissertation, Groningan University.*

Shelton B.R. & Searle C.L. The influence of vision on the absolute identification of sound-source position. 1980, *28(6)*, 589-596.

Spelke, E. S, Born, W. S., & Chu, F. Perception of moving sounding objects by four month old infants. *Perception.* 1983, *12*, 719-732.

Steiger, H, & Bregman, A. S. Capturing frequency components of glided tones: Frequency separation, orientation and alignment. *Perception and Psychophysics.* 1981, *30,* 425-435.

Stein, B. E, & Meredith, M. A. The merging of the senses. 1993. Cambridge MA: MIT Press.

Stein, B. E., Meredith, M. A., & Wallace, M. T. Development and neural basis of multisensory integration. 1994 In The development of intersensory perception. Comparative perspectives. Lewkowicz D. J., & Lickliter R. (Eds.)

Stein , B. E., Development of the superior colliculus. *Annual review of Neuroscience.* 1984, *7*, 95-125.

Streri, A, & Molina, M. Constraints on intermodal transfer between touch and vision in infancy.1994 In. The development of intersensory perception, Comparative perspectives. Lewkowicz D. J., & Lickliter R. (Eds.)

Streri, A., & Pêcheux, M, -G., Vision to touch and touch to vision transfer of form in 5-month-old infants. *British Journal of Developmental Psychology,* 1986, *4,* 161-167.

Sumby, W.G. & Pollack, I. Visual contributions to speech inteligibility in noise. *Journal of the Acoustical Society of America,* 1954,*26*, 212-215.

Summerfield, Q. Some preliminaries to a comprehensive account of audiovisual speech perception.Pp 3-52, Hearing by Eye: The Psychology of Lip-Reading. In Dodd & Campbell eds 1987 LEA.

Summerfield, Q. Lipreading and audiovisual speech perception. *Transactions of the Royal Society,* 1991, *B.*

Thurlow, W. R. , & Jack, C. E. Certain determinants of the "ventriloquism effect". *Perceptual and Motor Skills,* 1973, *36,* 1171-1184.

Van Noorden, L. P. A. S. Temporal Coherence in the Perception of Tone Sequences. 1975, *Unpublished Doctoral Dissertation.* Eindhoven University of Technology.

Walker, J. T., & Scott, K. J. Auditory-visual conflicts in the perceived duration of lights, tones and gaps. *JEP Human Perception and Pereormance,* 1981, *7,*1327-1339.

Walton, G.E., & Bower, T.G.R. Amodal representation of speech in infants. *Infant behaviour and development,* 1993, *16,* 233-243.

Warren D.H, McCarthy T.J & Welch R.B. Discrepancy and non-discrepancy methods of assessing visual-auditory interaction. *Perception and Psychophysics*, 1983, *33(5)*, 413-419.

Warren, D. H., & Pick, H. L., Jr. Intermodality relations in blind and sighted people. *Perception and Psychophysics,* 1970, *8,* 430-432.

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

Watanabe, J. & Iwai, E. Neuronal activity in visual , auditory and polysensory areas in the monkey temporal cortex during visual fixation task. *Brain Research Bulletin,* 1991, *26,* 583-592.

Watson C.S, & Mittler B.T. Time -Intensity equivolence in auditory lateralization: A graphical method. *Psychonomic Science*, 1965, *2,* 218-219.

Welch, R. B., & Warren, D. H. Immediate perceptual response to intersensory discrepency. *Psychological Bulletin*, 1980, *88,* 638-667.

Welch, R. B., DuttonHurt, L., D., & Warren, D. H. Contributions of audition and vision to temporal rate perception. *Perception and Psychophysics,* 1986, *4,* 294-300.

Wertheimer, M. Psychomotor co-ordination of auditory and visual space at birth, *Science*, 1961, *134,* 1692.

Wessel, D. L. Timbre space as a musical control structure. *Computer music journal,* 1979, *3(2),* 45-52.

Willey, C. F., Inglis, E. & Pearce, C. H. Reversal of auditory localization. *Journal of Experimental Psychology.* 1973, *14*, 577-580.

Withington-Wray, D. J., Binns, K. E., & Keating, M. J. The maturation of the superior collicular map of auditory space in the guinea pig is disrupted by developmental visual deprivation. *European Journal of Neuroscience,* 1990, *2,* 682-692.

Withington-Wray, D. J. & Keating, M. J. Visual or auditory deprivation of an electrophysical map of auditory space in the guinea pig. *Society for Neurosciences Abstracts,* 1989, *15*, 291

Yost W.A. Lateral position of sinusoids presented with interaural intensive and temporal differences, *Journal of the Acoustical Society of America,* 1981, *70(2),* 397-409.

Yost W.A. & Dye R.H. Discrimination of interaural differences of level as a function of frequency. *Journal of the Acoustical Society of America,* 1988, *83(5),* 1846-1851.

Yost W.A, Hafter E.R, Lateralisation. in Directional Hearing Yost & Gourevitch eds. 1987, Springer-Verlag.

Young, P. T. Auditorylocalization with acoustical transposition of the ears. *Journal of Experimental Psychology.* 1928, *11,* 399-429.