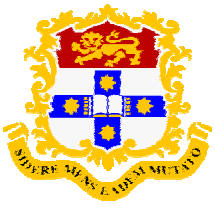
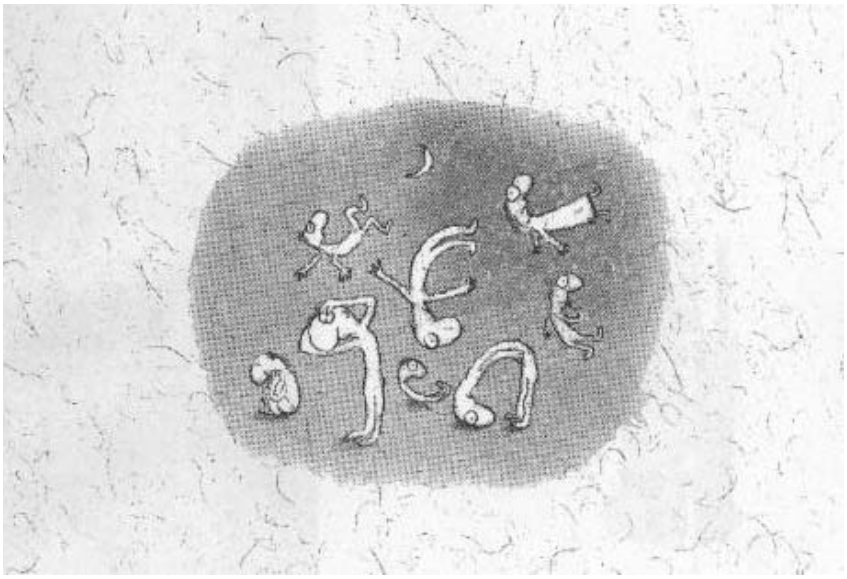
# Spatial Hearing with Simultaneous Sound Sources: A Psychophysical Investigation

A thesis submitted in fulfilment of the requirements for the degree of Doctor of Philosophy

Faculty of Medicine,
The University of Sydney.
April 2004.



by
Virginia Ann Best



God bless the lost, the confused, the unsure,
the bewildered, the puzzled, the mystified,
the baffled, and the perplexed.
Amen.

- Michael Leunig -

# Acknowledgements

I would first like to acknowledge and thank my supervisors, Simon Carlile and André van Schaik, for absolutely everything. Simon's enthusiasm, wisdom, and knack for story-telling first got me excited about ears, kept me inspired, and will stay with me always. I am grateful for his unfailing belief in me. The opportunities and responsibilities he gave me have made me so much stronger and I thank him for making me do things I didn't think I could do! André's contribution to this work (and my character) has been enormous. From the very start he excused my naivety and trusted my ability to understand and learn. He shared with me the brilliant things he was doing and, more importantly, thinking. Working with André has made me a more critical thinker, a clearer communicator, and a much better anagram unscrambler. Above everything, I treasure how effortlessly we blended work with fun and how often we read each others minds. André, heel hartelijk dank voor allehulp.

A most important aspect of this work was the amazing environment it took place in, and I most affectionately acknowledge all in the Auditory Neuroscience Lab. A special thanks must go to Johahn Leung, for taking me under his wing when I arrived in the lab, and for being there with friendship and support ever since. Jogi was the one who always had the answer I needed, but always made sure I worked for it so that I would learn something. My warmest thanks also to Craig Jin for his ideas, his advice, and his unrivalled ability to get things moving. Craig's fascination with everything took this work down many important avenues, and his famous suspicions averted many disasters! I also want to thank Ben Dickson, whom I felt I had known for years as soon as he arrived in the lab. It was wonderful to share an office with him, as well as the sometimes rocky PhD road, some good and bad music, and a lot of laughs.

My friends are the ones who have always kept me balanced and smiling, and I thank them all wholeheartedly. I am especially grateful to them for accepting those periods of distraction where I completely forgot how to be fun. A (fully) special thank you to Sarah-Jane, for a million things that I needn't mention because she just *knows*. And to Haydn, who faced the writing-up stage with me, for inspiring me to finish through his determination and energy, his love and care, and that great big smile. Finally, I would like to acknowledge my beautiful family for their undemanding love and quiet encouragement. I am truly the luckiest person in the world to have them. I dedicate this thesis to them, but want to assure them that they are not obliged to read past this page! xo

# Abstract

This thesis provides an overview of work conducted to investigate human spatial hearing in situations involving multiple concurrent sound sources. Much is known about spatial hearing with single sound sources, including the acoustic cues to source location and the accuracy of localisation under different conditions. However, more recently interest has grown in the behaviour of listeners in more complex environments. Concurrent sound sources pose a particularly difficult problem for the auditory system, as their identities and locations must be extracted from a common set of sensory receptors and shared computational machinery. It is clear that humans have a rich perception of their auditory world, but just how concurrent sounds are processed, and how accurately, are issues that are poorly understood. This work attempts to fill a gap in our understanding by systematically examining spatial resolution with multiple sound sources.

A series of psychophysical experiments was conducted on listeners with normal hearing to measure performance in spatial localisation and discrimination tasks involving more than one source. The general approach was to present sources that overlapped in *both frequency and time* in order to observe performance in the most challenging of situations. Furthermore, the role of two primary sets of location cues in concurrent source listening was probed by examining performance in different spatial dimensions. The binaural cues arise due to the separation of the two ears, and provide information about the lateral position of sound sources. The spectral cues result from location-dependent filtering by the head and pinnae, and allow vertical and front-rear auditory discrimination.

Two sets of experiments are described that employed relatively simple broadband noise stimuli. In the first of these, two-point discrimination thresholds were measured using simultaneous noise bursts. It was found that the pair could be resolved only if a binaural difference was present; spectral cues did not appear to be sufficient. In the second set of experiments, the two stimuli were made distinguishable on the basis of their temporal envelopes, and the localisation of a designated target source was directly examined. Remarkably robust localisation was observed, despite the simultaneous masker, and both binaural and spectral cues appeared to be of use in this case. Small but persistent errors were observed, which in the lateral dimension represented a systematic shift *away* from the location of the masker. The errors can be explained by interference in the processing of the different location cues. Overall these experiments demonstrated that the spatial perception of concurrent sound sources is highly dependent on stimulus characteristics and configurations. This suggests that the underlying spatial representations are limited by the accuracy with which acoustic spatial cues can be extracted from a mixed signal.

Three sets of experiments are then described that examined spatial performance with speech, a complex natural sound. The first measured how well speech is localised in isolation. This work demonstrated that speech contains high-frequency energy that is essential for accurate three-dimensional localisation. In the second set of experiments, spatial resolution for concurrent monosyllabic words was examined using similar approaches to those used for the concurrent noise experiments. It was found that resolution for concurrent speech stimuli was similar to resolution for concurrent noise stimuli. Importantly, listeners were limited in their ability to concurrently process the location-dependent spectral cues associated with two brief

speech sources. In the final set of experiments, the role of spatial hearing was examined in a more relevant setting containing concurrent streams of sentence speech. It has long been known that binaural differences can aid segregation and enhance selective attention in such situations. The results presented here confirmed this finding and extended it to show that the spectral cues associated with different locations can also contribute.

As a whole, this work provides an in-depth examination of spatial performance in concurrent source situations and delineates some of the limitations of this process. In general, spatial accuracy with concurrent sources is poorer than with single sound sources, as both binaural and spectral cues are subject to interference. Nonetheless, binaural cues are quite robust for representing concurrent source locations, and spectral cues can enhance spatial listening in many situations. The findings also highlight the intricate relationship that exists between spatial hearing, auditory object processing, and the allocation of attention in complex environments.

# Declaration

This thesis describes original work carried out in the Department of Physiology at the University of Sydney. I certify that all material in this thesis which is not my own work has been identified and that no material has previously been submitted and approved for the award of a degree by this or any other University.

Virginia Best
April 2004

Portions of this work have appeared in the following publications:

V. Best, A. van Schaik, C. Jin and S. Carlile (2004). "Sharing auditory space: Sound localisation in the presence of a concurrent masker." Acta Acustica united with Acustica (special issue on Spatial and Binaural Hearing) (submitted).

V. Best, A. van Schaik and S. Carlile (2004). "Separation of concurrent broadband sound sources by human listeners." Journal of the Acoustical Society of America 115(1):324-336.

V. Best, A. van Schaik and S. Carlile (2003). "Two-point discrimination in auditory displays." Proceedings of the 9th International Conference on Auditory Display, Boston, USA, pp. 17-20.

V. Best, A. van Schaik and S. Carlile (2003). "Spatial effects on the segregation of sounds in virtual auditory space." Proceedings of the 8th Western Pacific Acoustics Conference, Melbourne, Australia, p. 1090M.

V. Best, A. van Schaik and S. Carlile (2002). "The perception of multiple broadband noise sources presented concurrently in virtual auditory space." Proceedings of the Audio Engineering Society 112th Convention, Munich, Germany.

C. Jin, V. Best, S. Carlile, T. Baer and B. Moore (2002). "Speech Localization." Proceedings of the Audio Engineering Society 112th Convention, Munich, Germany.

# Table of contents

# Chapter 1: Sound localisation

## 1.1 Overview

This chapter is divided into four major sections. In the first section, an overview of the auditory system is given, including its basic organisation and coding strategies. This is provided as a background to the second section which discusses auditory spatial processing. Here the acoustic cues to sound source location are described and their encoding by the nervous system is considered. These sections are intended only as background. Excellent comprehensive reviews of these areas are available, and are referred to as appropriate in the text. In the third section, a review of some of the literature concerning human auditory spatial performance is presented. This review is aimed at painting a picture of the level of accuracy with which listeners localise and discriminate *single* sound sources. In the final section, the notion of spatial resolution for *concurrent* sound sources is considered, and the aims of the thesis are introduced.

## 1.2 Organisation of the auditory system

### 1.2.1 The ear

An illustration of the human ear is provided in Figure 1.1. The outer ear consists of the pinna (a cartilaginous flap) surrounding a resonant cavity called the concha, which leads to the eardrum via the ear canal. The outer ear acts as an acoustic filter and transforms incoming soundwaves, contributing to the frequency spectrum of the pressure wave that ultimately vibrates the eardrum. This energy is transferred mechanically by the middle ear to the cochlea. The middle ear is an air-filled cavity that contains three small bones (called the ossicles) connected in a chain. The action of the middle ear serves in impedance matching: increasing the efficiency with which air-borne sound energy is transferred to the fluid of the cochlea. The cochlea is essentially a coiled tube, located in the temporal bone of the skull, which is divided

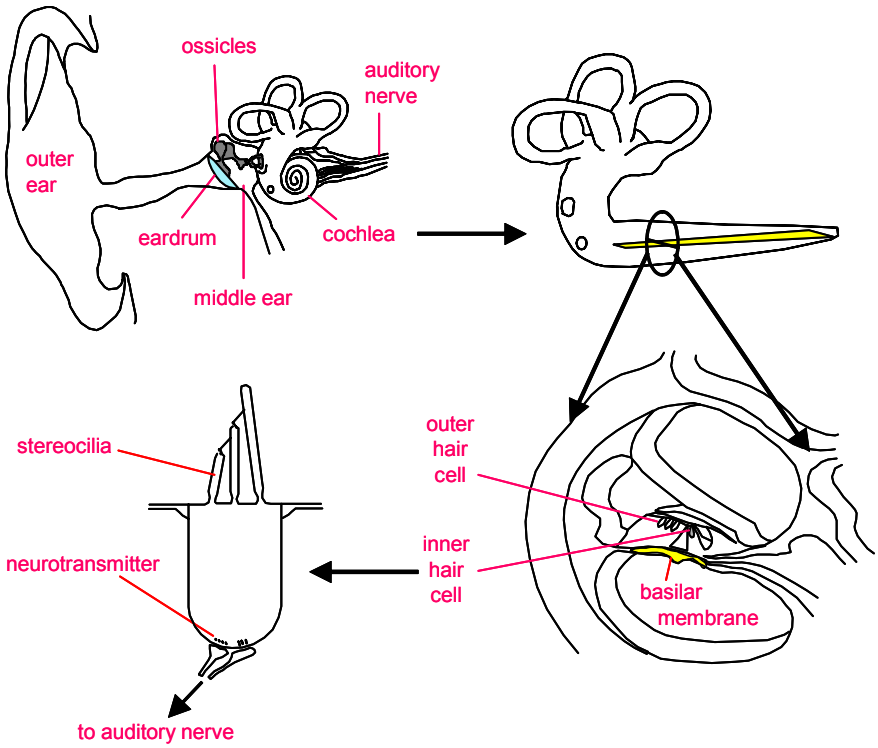



**Figure 1.1** The human ear. The outer ear filters incoming sounds before they reach and vibrate the eardrum. This energy is transferred mechanically via the middle ear ossicles to the fluid of the cochlea. The cochlea is a coiled tube that is divided along its length by membranes into three fluid-filled compartments. The basilar membrane vibrates within the fluid, and its motion is converted to neural signals by the hair cells. *Picture adapted with permission from André van Schaik.*

along its length by membranes into three fluid-filled compartments. The vibration of the eardrum causes pressure waves to travel through the fluid of the cochlea, setting up travelling waves in the lower basilar membrane, which is approximately 35 mm long.

It is along this membrane that the sensory receptors of the auditory system reside. The outer and inner hair cells are arranged in rows along the length of the membrane. The inner hair cells are the primary mechano-electric transducers of the system, converting the motion of the basilar membrane into neural signals. This mechanism involves the deflection of stereocilia located on the apical ends of the hair cells: this movement is thought to influence mechano-sensitive ion channels in the stereocilia membrane, causing voltage fluctuations within the cell. Depolarisation leads to neurotransmitter release from the basal ends of the hair cells, which contact fibres of the auditory nerve (part of the vestibulocochlear or 8th cranial nerve). The

auditory nerve serves as the primary transmission line to (and from) the brain. The outer hair cells also transduce stereocilia displacement, but their primary action is to generate positive feedback forces that enhance the motion of the basilar membrane. The outer hair cells are also subject to efferent innervation from the central auditory pathways. For extensive detail on the structures and mechanisms introduced here, see Pickles (1988).

### 1.2.2 Basic coding strategies

Perhaps the most important organisational feature of the auditory system is its tonotopicity, or parallel frequency arrangement. Input sounds are essentially decomposed into their component frequencies in the cochlea. The basilar membrane varies smoothly in mass and stiffness from base to apex and, as a result, a particular frequency will maximally vibrate the membrane at a characteristic position along its length. Thus energy at different frequencies is physically separated at this point. Auditory nerve fibres contacting the cochlea then carry frequency-specific information to the brainstem whilst retaining their relative positions. This results in a 'place code' where the location of neural activity in the auditory nerve indicates the frequency spectrum of the sensory input. Importantly, this organisation is preserved at all levels along the pathways ascending (and descending) through the auditory system.

Information about a sound stimulus is also available in the firing rates and firing patterns of auditory neurons. In general, stimulus intensity is coded via firing rate, where an increase in the intensity of stimulation at a particular frequency will be relayed as an increase in the firing rate of neurons tuned to that frequency. Furthermore, the pattern of firing carries information about the stimulus waveform. For low frequency stimuli (below about 4-5 kHz), neurons tend to fire at a particular phase of the waveform each cycle – a process called 'phase locking' – and thus the time between spikes is related to the period of the stimulating waveform. Phase locking, and temporal firing patterns in general, are particularly important for coding temporal cues to sound source location (see below).

A final important feature of auditory coding, and neural coding generally, is that information is represented by *ensembles* of neurons. Take as an example the motion of the basilar membrane elicited by a pure tone stimulus. This will result in the

firing of hair cells at the position of maximum displacement, as well as a weaker firing of hair cells in surrounding regions. As each inner hair cell is contacted by several auditory nerve fibres (about 20, according to Pickles, 1988), then a clustered population of neurons in the auditory nerve will be active. The activity will display a graded profile across the tonotopic array with a peak corresponding to the frequency of the tone. This notion of population coding holds in all parts of the auditory system.

### 1.2.3 Central auditory pathways

The anatomy of the auditory nervous system is extremely complex. The many pathways display a high degree of convergence and divergence, and contain a relatively large number of sub-cortical nuclei compared to other sensory systems. A very brief description of the major structures and pathways is given in the following sections, and Figure 1.2 provides an illustration. For more information, excellent chapters are available on the anatomy and physiology of these structures in Popper and Fay (1992) and Webster *et al.* (1992).

Afferent information from the inner hair cells is carried in type I spiral ganglion neurons, travelling in the auditory part of the vestibulocochlear nerve. Also travelling in this nerve, type II spiral ganglion neurons carry afferent information from the outer hair cells. The basal ends of the outer hair cells and the afferent fibres of the inner hair cells also receive inputs from *efferent* fibres travelling in the olivocochlear bundle. Descending information is thought to influence not only the cochlea but every nucleus in the auditory pathway, and has been implicated in a huge variety of functions including frequency selectivity, intensity coding, and selective attention. This efferent pathway will not be addressed here, but reviews are available elsewhere (see Huffman and Henson, 1990; Spangler and Warr, 1991; and a recent update in Suga and Ma, 2003).

All afferent spiral ganglion cells synapse in the cochlear nucleus of the brainstem, making contact with its three divisions: the anterior ventral cochlear nucleus (AVCN), the posterior ventral cochlear nucleus (PVCN) and the dorsal cochlear nucleus (DCN).

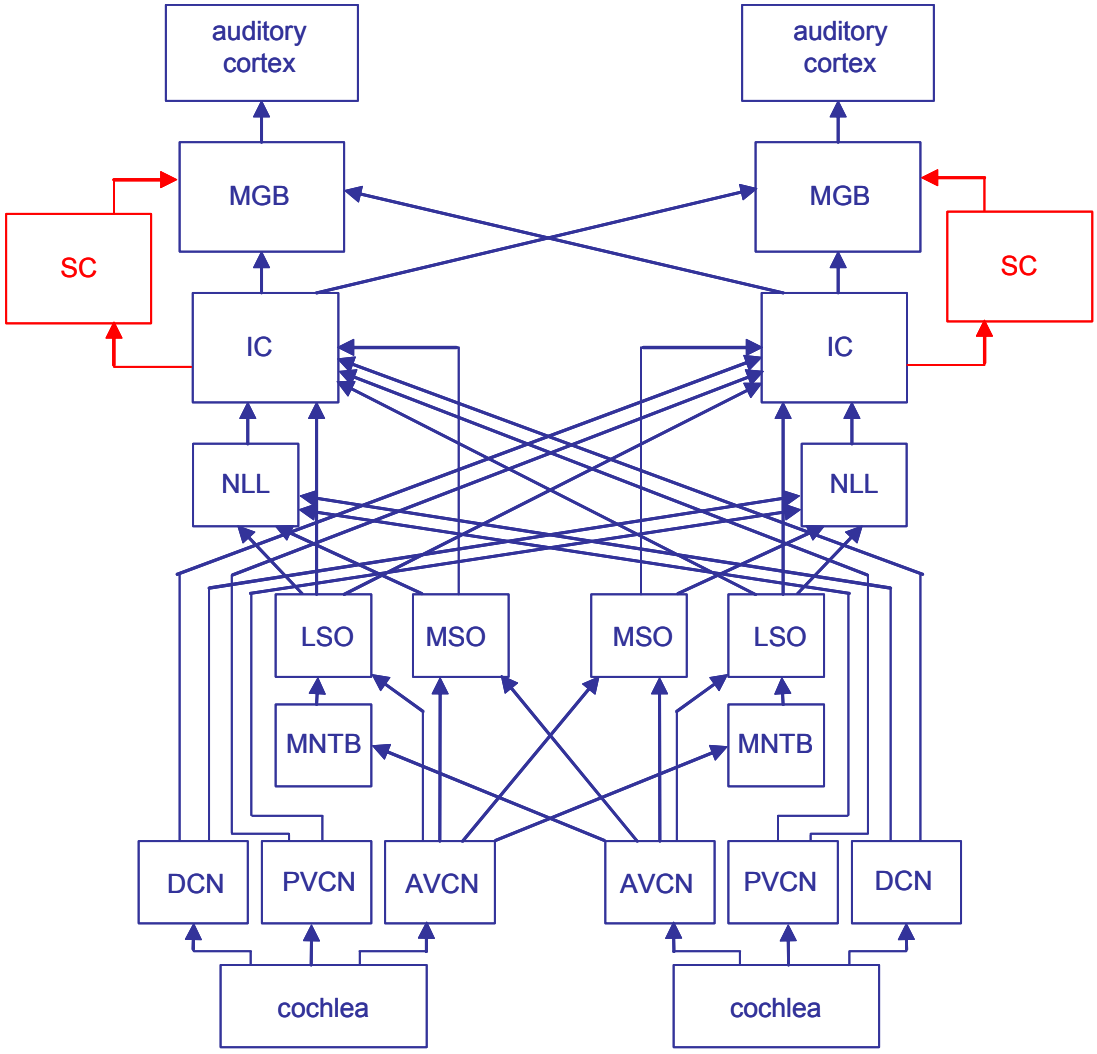



**Figure 1.2** The major ascending pathways of the auditory system. This representation does not include commissural connections between left and right nuclei at the same level, subdivisions of the various nuclei, or any of the descending pathways. The superior colliculus (SC, shown in red) is not strictly an auditory nucleus but is relevant to discussions of auditory spatial processing. *This "crude summary" adapted with permission from an original by Professor David C. Mountain (available at http://earlab.bu.edu/intro/Auditorypathways.aspx).*

In the pons, the superior olivary complex (SOC) receives convergent inputs from the left and right cochlear nuclear complexes. The bilateral innervation of this complex is known to be essential for sound localisation; when crossing fibres of the trapezoid body are cut, localisation behaviour is severely disrupted (Moore *et al.*, 1974). Two major 'spatial' nuclei are found in the SOC. The medial superior olive (MSO) receives excitatory inputs from both left and right AVCN and is known to be involved in sound localisation by encoding interaural time differences (see below). The smaller lateral superior olive (LSO) also receives bilateral inputs, an excitatory

input from the ipsilateral AVCN, and an inhibitory input from the contralateral AVCN (via the ipsilateral medial nucleus of the trapezoid body, MNTB). This arrangement renders the cells of this structure sensitive to interaural differences of intensity, which are also important in sound localisation (see below).

The lateral lemniscus is a major auditory tract carrying axons from the AVCN, DCN, PVCN, MSO and LSO up to the inferior colliculus (IC). Here there is some crossover of fibres between the left and right nuclei, but the majority of IC axons form the ipsilateral brachium of the IC. This tract leaves the midbrain and terminates in the medial geniculate body (MGB) of the thalamus. The main projection of the MGB is to primary auditory cortex, and there are extensive projections back to the MGB from cortex. However, the various divisions of the MGB also communicate with a range of other brain areas. Inputs arise from the brainstem reticular formation, other thalamic nuclei, vestibular nuclei, spinal cord nuclei and superior colliculus. Target structures include auditory and non-auditory cortices and the amygdala. So the MGB is an important relay structure and is implicated in attention and multi-sensory functioning.

The superior colliculus, while primarily a visual reflex centre, receives highly organised auditory inputs from the inferior colliculus. Here spatial auditory information is arranged in alignment with visual and motor spatial maps (Oliver and Huerta, 1992) and is important in the triggering of eye and head movements (Harris *et al.*, 1980; Wurtz and Albano, 1980).

The primary auditory cortex is found in the temporal lobe. Many structural and functional organisations have been described for this area, and it is here that complex features of auditory objects are represented. The importance of the primary auditory cortex in spatial hearing is indicated by the inability of mammals to localise sound sources in the contralateral hemifield after lesions in this area (Kavanagh and Kelly, 1987). Primary auditory cortex makes connections with several other important cortical areas, including Wernicke's area (speech reception) and Broca's area (speech production and language).

## 1.3 Auditory spatial processing

Let us now consider how the spatial qualities of an object are encoded. Spatial location is not represented on the sensory epithelium in the auditory modality, but

must be computed from the acoustic input to the ears. The first class of cue is the *binaural* cues that arise as a result of the separation of the two ears in space. Two main binaural cues define the location of a sound source relative to a listener, and these are the interaural differences in time and level (ITD and ITD). Further to these, *monaural* cues are available from the effect of the auditory periphery on incoming sound waves.

These cues are considered in turn in the following sections in terms of their physical characteristics, their contribution to human sound localisation, and their neural encoding by the structures of the auditory system. Note that these issues are covered in far greater depth elsewhere (e.g. Blauert, 1983; Middlebrooks and Green, 1991; Carlile, 1996b; Popper and Fay, 1992).

Much of our insight into the cues to sound source location has come from neurophysiological and psychoacoustic studies. In these studies, acoustic stimuli are presented in a range of ways (to human or animal subjects). In the most realistic stimulation paradigms, sounds are presented via loudspeakers in external space (referred to here as 'free-field' listening). However for a more controlled environment, many researchers have opted for headphone presentation. In this case, the percept is of a sound that is located inside the head, but binaural cues can be varied to give the impression of lateral displacement within the head, known as lateralisation. More recently, the development of virtual auditory space (VAS) technology has greatly expanded the scope of research in this area. Using this approach, realistic spatial cues can be incorporated into headphone listening to give an accurate simulation of the free-field listening experience (see Chapter 2).

### 1.3.1 Interaural time differences

#### Characteristics of the ITD

ITDs arise as a result of the separation of the two ears in space. For a sound source located off the midline, the wavefront must travel a greater distance to reach the far ear compared to the near ear, setting up a time delay between the two ears (Figure 1.3). This time delay ranges from 0 for a sound on the median plane, to an average of 690 μs for a source located directly to the side, with values varying sinusoidally with location. The exact value of the ITD for a particular direction is dependent primarily on the head-width of the listener, but is also slightly frequency-dependent (see Kuhn,

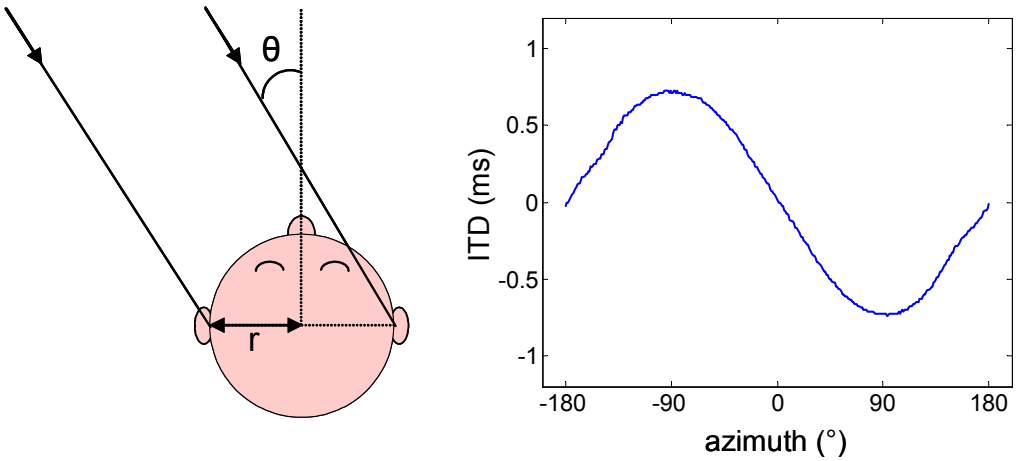



**Figure 1.3** Left: Assuming a spherical head of radius r, the path length difference d to the two ears for a distant source at an angle of θ is given by d = r θ + r sin θ and the equivalent ITD is d / c where c = speed of sound (343m/s). Right: The ITD as a function of azimuth for a human subject (the author). The ITD was calculated by cross-correlation of the left and right ear impulse responses (bandpass filtered between 300 and 1000 Hz). Note that this is an overall measure that conceals small frequency-dependencies of the ITD.

1987).

The ITD has two main components: an onset disparity and an ongoing disparity. For all sounds, including very transient sounds such as clicks, there exists an onset ITD that comes from the disparity between the arrival of a sound at each ear (note there is an equivalent disparity at the offset). This cue is available across all frequencies. For continuous sounds there is also an ongoing ITD between the signals at the two ears. For pure tones, this emerges a phase difference in the fine structure of the signals and is a particularly reliable cue for low frequencies (below about 1.5 kHz). For higher-frequency tones, however, the interaural phase difference becomes an ambiguous cue to location as the period becomes much shorter than the interaural delays. For complex waveforms, however, it is known that interaural delays in the slowly varying envelope of high-frequency channels (above 1.5 kHz) can be processed by the auditory system (Henning, 1974, 1980; McFadden and Pasanen, 1976).

In terms of the relative potency of the different ITD cues, headphone studies have shown that ongoing ITD is a more potent indicator of lateral position than is onset/offset ITD. The latter appears to only have a strong influence when the signal is brief (shorter than about 10 ms: Tobias and Schubert, 1959) or the ongoing ITD is

ambiguous (Kunov and Abel, 1981; Buell *et al.*, 1991). Furthermore, there is some evidence that the high-frequency ongoing ITD is a weaker cue than the low-frequency ongoing ITD for lateralisation (Yost, 1976; Bernstein and Trahiotis, 1982).

It is commonly reported that the ITD is the most potent of the cues to sound source location, and there is certainly good evidence in favour of this idea from experiments carried out under anechoic conditions. Wightman and Kistler (1992) had subjects localise sound stimuli presented over headphones in virtual auditory space. They fixed the ITD to a value corresponding to a location directly to the right of the listener, but kept the overall ILD and the spectrum at each ear varying as usual for different locations. They found that horizontal estimates were strongly clustered at the ITD location. This suggests that in particular circumstances, the ITD can override the other localisation cues if they are conflicting. This might not be the case, however, in echoic conditions where the consistency of the ITD cue is severely reduced by reflections and reverberation (see for e.g. Shinn-Cunningham and Kawakyu, 2003).

**Neural coding of the ITD**

In 1948, Jeffress proposed a theoretical model of how interaural delays may be transformed into a neural representation of lateral position (Jeffress, 1948). His elegant model consisted of an array of 'coincidence detectors': neurons that only fired when input was received simultaneously from the left and right ears. To encode an ITD of zero (for a stimulus equidistant to the two ears), the left and right input signals must simply be kept in synchrony during transmission to the array. For one of these units to respond preferentially to a non-zero ITD, however, the delay between the two ears must be compensated for during transmission to the array in order for the signals to arrive simultaneously. Jeffress proposed that appropriate neural delays could be established by altering the relative path lengths of the inputs in order to alter their relative travel times. The MSO is perfectly suited to encode ITDs in this fashion, as it receives temporally accurate signals from the two ears via the AVCN (Pickles, 1988). In fact it has been shown that the bushy cells of the AVCN are *much more* precisely locked to the stimulus waveform than auditory nerve fibres (Joris *et al.*, 1994). Furthermore, the majority of cells in the MSO are EE (excitatory-excitatory) and are most responsive to low frequencies, where phase information is preserved. It has been confirmed physiologically in several mammals that MSO neurons are sensitive to ITD and that the "best ITD" of a neuron can be predicted from its response delays to

monaural inputs from the two ears (Yin and Chan, 1990; Goldberg and Brown, 1969; Spitzer and Semple, 1995).

However, anatomical support for the Jeffress model has not been strong. While delay-line type organisations have been reported in the brainstem of birds (Carr and Konishi, 1990; Overholt *et al.*, 1992), only scant evidence is available in mammals (Smith *et al.*, 1993). Furthermore, best ITDs of neurons in mammalian MSO are often outside the range of ITD that is physically relevant to the animal (Brand *et al.*, 2002; this is also true for IC: McAlpine *et al.*, 2001; and cortex: Fitzpatrick *et al.*, 2000). It appears likely that the *dynamic range* rather than the *maximum response* of ITD functions is the key to ITD encoding. Recent models of ITD processing are based on this idea. McAlpine *et al*. (2001) proposed that lateral position may be determined by the relative activity in two broadly-tuned spatial channels, with activity varying along the slope of the ITD functions. In an important recent study, Brand *et al*. (2002) showed that contralateral inhibition is crucial for the tuning of the slopes of ITD functions to relevant ITD ranges. In their scheme, precise timing of this inhibition allows differential sensitivity to ITD, without the need for precise anatomical delay lines (this is an appealing concept as the auditory system is specialised for temporal coding).

So although there are continuing developments in our understanding of ITD encoding, it is agreed that the brain performs a comparison of inputs to the two ears and is exquisitely sensitive to interaural delays. This functionality is commonly incorporated into computational auditory models by means of interaural cross-correlation. Several such models have played important roles in extending our understanding of various binaural phenomena (see Colburn, 1996 for review)

### 1.3.2 Interaural level differences

**Characteristics of the ILD**

ILDs arise primarily due to the fact that the head separates the two ears, and acts as an acoustic obstacle. The result of this is that a sound originating from one side will be more intense in the near ear as compared to the far ear, which is ‘shadowed’ (Figure 1.4). Also contributing to this effect, the pinnae are directional in many animals

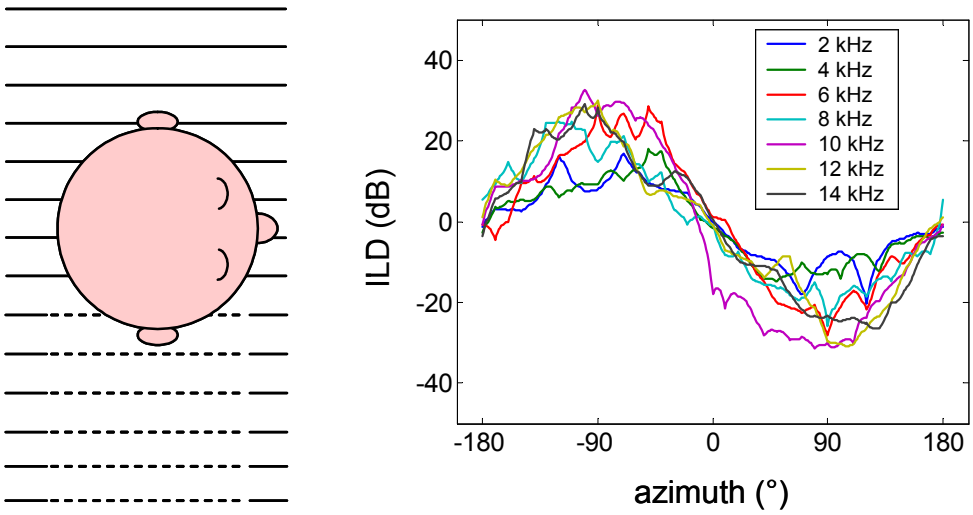



**Figure 1.4** Left: The head casts an 'acoustic shadow' that results in an interaural level difference. Right: The ILD as a function of azimuth for a human subject (the author). The ILD was calculated by subtracting the magnitude response of the right impulse response from the left at the frequencies specified in the legend. The frequency-dependence is attributable to differential effects of the head, shoulders and pinnae.

including humans, and the pinna of the near ear acts to amplify high frequency sounds (Shaw, 1974). Importantly, the ILD is frequency-dependent as a result of differential filtering effects of the head and pinnae (Figure 1.4). In general however, the ILD increases with horizontal displacement from the front to the side. This monotonic variation is similar to that of the ITD, where the rate of change is greatest in the frontal region. Values of ILD range from 0 dB on the median plane to as large as 20 dB at the side. The ILD, however, is negligible at low frequencies (below about 1.8 kHz) where the wavelength is long enough to diffract around the head and not be influenced by the small structures of the pinnae.

The above observations are based on sound sources positioned at a reasonable distance from the head of a listener (1m or greater). Recent work has shown that for nearer sources, ILDs vary far more dramatically and even occur at low frequencies (Brungart and Rabinowitz, 1999; Shinn-Cunningham *et al.*, 2000).

### Neural coding of the ILD

The main class of neurons that are associated with the coding of ILDs receive an excitatory input from one ear and an inhibitory input from the other. These are termed EI (excitatory-inhibitory) or IE (inhibitory-excitatory) neurons (Irvine, 1992). EI cells

are concentrated in the LSO, and this is thought to be the initial site of ILD encoding (Tollin, 2003). Here it has been shown that high-frequency neurons are relatively over-represented. Furthermore, a topographic representation of ILD exists within frequency bands, suggesting that ILD is encoded on a spectral basis (Irvine, 1992). EI neurons in the LSO respond according to the balance of ipsilateral excitation and contralateral inhibition, and hence are sensitive to imbalances in intensity between the left and right ear inputs. Their discharge rate varies sigmoidally with ILD.

Delgutte *et al.* (1995) showed that ILDs were the most potent of the localisation cues for the directional sensitivity of high-characteristic-frequency neurons in the cat IC. The IC receives inputs from the contralateral LSO, and hence at this level (and above) it is IE cells that maintain a representation of ILD. These high-frequency ILD sensitive neurons also show sensitivity to onset time differences and envelope delays in complex sounds (Yin *et al.*, 1984; Caird and Klinke, 1987).

### 1.3.3 Spectral cues

**The cone of confusion**

The binaural system provides excellent directional information by making comparisons between the two samples of a sound it receives. However, as the receivers (the ears) are placed roughly symmetrically on either side of the head, the system is inherently ambiguous in the coding of three-dimensional space. For example, there are a number of locations that can give rise to a particular ITD. In fact, an ITD defines a set of locations on a rough cone centred on the interaural axis (Figure 1.5). A similar set of locations are defined by a particular ILD, although the shape is less regular due to the differential effects of the pinnae. This cone of equal binaural value was titled the “cone of confusion” by Woodworth (1938) and was described in detail by Mills (1972).

These ambiguities may in some instances be resolved by head movements (Wallach, 1940; Wightman and Kistler, 1999) but even for extremely brief signals where no head movements are possible, and under virtual stimulation where the spatial cues are fixed, listeners are able to resolve locations on the cone of confusion. The explanation for this is that the human auditory system has *monaural* spatial cues available to it (reviewed in Blauert, 1983).

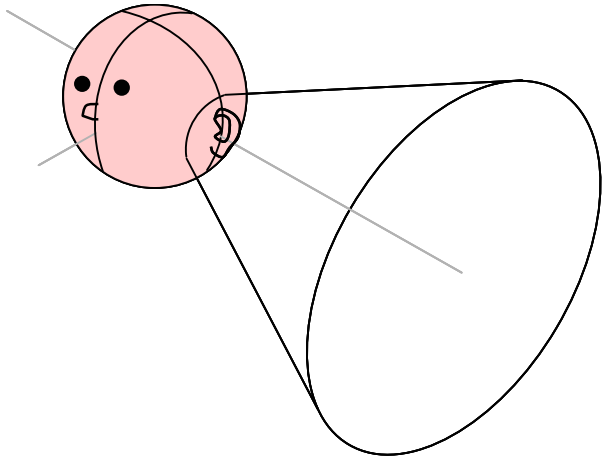


**Figure 1.5** A cone of confusion. The surface of the cone approximates the locus of sources delivering the same binaural cues. The cone is derived from the analysis of ITD on the horizontal plane (Figure 1.3), extended to include elevation also. Assuming a spherical head of radius r, the path length difference d to the two ears for a distant source at an azimuth of θ and an elevation of ε is given by $d = r ( \sin^{-1} (\sin \theta \cos \varepsilon ) + \sin \theta \cos \varepsilon)$ and the equivalent ITD is d / c where c = speed of sound (343m/s).

Sounds do not reach the ear canal directly, but are acoustically filtered by the auditory periphery. This includes diffraction by the head, shoulders and pinnae and resonance within the concha, resulting in a complex pattern of filtering that is highly individualised. Most importantly, the three-dimensional asymmetry of the auditory periphery means that the acoustic filtering provided is dependent on the *direction* of the incoming sound. For example, the pinna is directed somewhat forwards and hence interacts differently with sounds coming from the front or from the back. This provides a means for distinguishing front from back that is *not* provided by the binaural cues.

### Characteristics of the spectral cues

The spectral cues provided by the auditory periphery are commonly described by means of the head-related transfer function (HRTF). Several studies have examined the relationship between sound source location and the shape of the HRTF in humans and other animals. Importantly, the human HRTF has features across a large range of frequencies, and it has been suggested that spectral cue localisation requires the 4-16 kHz range (Hebrank and Wright, 1974). Acoustic studies have examined both the

'broadband shape' of these functions as well as particular features in the HRTF. While most of the features of the HRTF are location-dependent, there are some components (such as certain concha and ear-canal resonances) that do not vary with the direction of a stimulus, and thus in some studies of the spatial character of the spectral cues, these are removed to leave only the directional transfer function, or DTF (Middlebrooks and Green, 1990). Figure 1.6 shows an example of how the DTF changes with location. These DTFs were measured for one subject in anechoic space with the stimulus situated 1 m from the centre of the head (see section 2.5.2 for details of the recording procedure). It is clear that the human DTF changes with horizontal location, as seen in Figure 1.6a (see also Carlile and Pralong, 1994). However, vertical planes have been studied most extensively in terms of DTF information, particularly the median vertical plane (MVP; the plane dividing the head symmetrically in half). This is because the MVP is the plane with the least binaural information (ITD and ILD are always equal to zero, assuming symmetry) and because there is only a single spectral cue available to the system (the spectral cue at each ear is approximately the same). Marked variations in the DTF can be seen along this plane (Figure 1.6b). It is important to note that the head and ears are *not* perfectly symmetrical in most people, and thus the MVP generally only approximates a plane of symmetry.

Different structures in the auditory periphery give rise to different spectral features in different parts of the audible spectrum. In general, the size of the structure defines the wavelengths with which it will interact and hence the frequencies that it amplifies or attenuates. Thus, low-frequency features are typically attributed to the head and shoulders, whilst detailed high-frequency features are attributed to the pinnae. Several authors have attempted to assign regions of the spectrum to dimensions in space along which they are most useful for localisation. It seems that the mid-frequency region is particularly important for elevation localisation (5.7 - 11.3 kHz according to Langendijk and Bronkhorst, 2002) and on the MVP a consistent notch in this region moves in frequency with changes in elevation (Hebrank and Wright, 1974; Butler and Belendiuk, 1977). Algazi *et al.* (2001) claimed that features below 3 kHz were also useful for off-midline elevation judgements (presumably due to head and torso influences). Relatively high frequency features (above 8 kHz) are thought to be most pertinent for front-back discrimination (Langendijk and Bronkhorst, 2002; Hebrank and Wright, 1974).

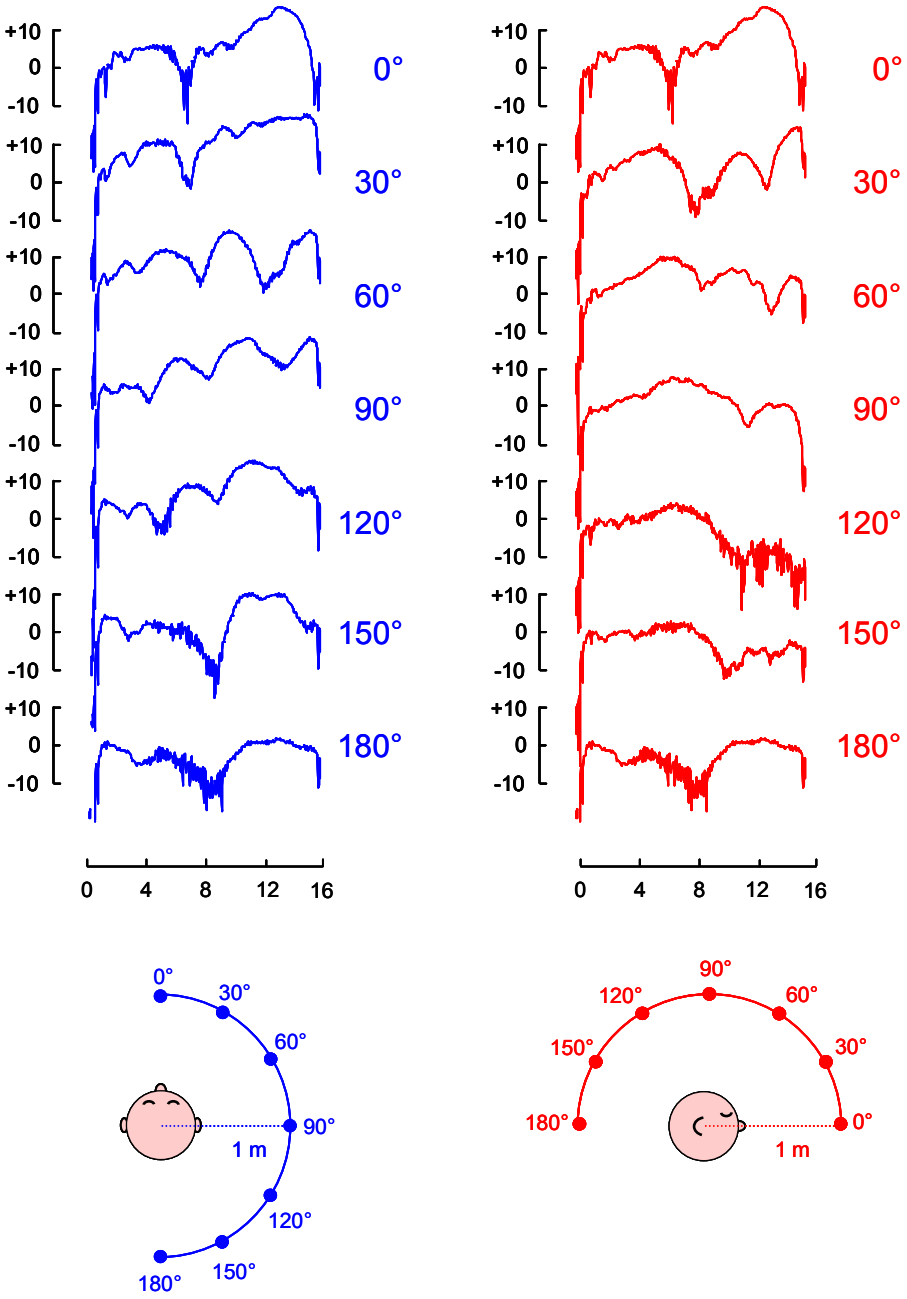



**Figure 1.6** Measured head-related transfer functions for a human subject (the author). Left: HRTFs for the right ear for 7 locations on the horizontal plane at the level of the ears. Right: HRTFs for the right ear for 7 locations on the median vertical plane.

Although the vertical plane (particularly the MVP) and the front-back dimension have been the major focus of research in this area, it is now generally agreed that the spectral cues are important for resolving locations anywhere around a cone of confusion.

## Contribution of the spectral cues to localisation

*Two spectral cues*

An important step in our understanding of sound localisation came from early studies that showed unequivocally that the pinnae are important (Angell and Fite, 1901; Roffler and Butler, 1967; Gardner and Gardner, 1973). However, as a different spectral cue is produced at each ear, the relative contribution of the two spectral cues and their possible interaction has been an ongoing question of interest.

Humanski and Butler (1988) demonstrated that the monaural spectral cue in the near ear contributes primarily to localisation in a given sagittal plane, but that the far ear contribution was also significant. Morimoto (2001) confirmed this notion, and by blocking each pinna cavity in turn with a mould and examining broadband noise localisation. He found that the near ear is increasingly important as a sound approaches its side, and that the far ear contribution is negligible at locations greater than 60° from the front. Hofman and Van Opstal (2003) extended this idea but concentrated on locations closer to the origin (+/- 30° azimuth and elevation) where binaural effects are likely to be maximal. They determined that spectral shape cues from the two ears are weighted to construct a percept of elevation, and the weighting depends in a gradual way on perceived azimuth.

In a recent experiment, Morimoto and colleagues tested the idea that spectral cues around *any* cone of confusion are similar (Morimoto *et al.*, 2003). They presented virtual stimuli created by giving MVP HRTFs different binaural cues (ITDs and ILDs), and measured localisation performance. They reported that the technique was quite successful for controlling the location of the sound image, in that their three subjects localised reasonably well. However they did not provide control localisation data using appropriate spectral cues, and so it is difficult to draw strong conclusions from their results.

These studies have all focussed on the contribution of the spectral cues to localisation within a sagittal plane (or cone of confusion), with the assumption that they do not contribute to the perception of lateral angle. This assumption has been verified recently by Macpherson and Middlebrooks (2002), who examined the weighting of various cues in lateral angle localisation and found the contribution of the near-ear spectrum to be negligible in the presence of binaural cues.

*Explaining narrow-band sound localisation*

Perhaps the best insight into how spectral cues are analysed has come from studies of narrow-band sound localisation. It has long been observed that narrow-band localisation in the MVP depends more on centre frequency than it does on *actual target location* (e.g. Pratt, 1930). A corollary of this is that narrow-band stimuli and simple noise stimuli with a prominent peak can give the percept of location (giving rise to the term "directional frequency bands": Blauert, 1983).

Several researchers have thought about this phenomenon in terms of the HRTFs, and several mechanisms have been considered. A popular conception is that the brain has stored HRTF 'templates' with which it compares the incoming spectrum to find the best match (Middlebrooks, 1992). Thus, a narrow-band stimulus would be localised at a location (on the appropriate cone of confusion) at which the centre frequency receives a large amount of gain from filtering by the pinna *relative to other frequencies*. For example, a narrow-band stimulus centred at 10 kHz might be localised at a location for which the HRTF has a prominent peak at 10 kHz. A more robust explanation however was originally proposed by Humanski and Butler (1988). In this model, narrow-band sounds are localised to the location (on the appropriate cone of confusion) that gives maximum gain to that frequency *relative to other locations*. Recently, Jin restated this idea and proposed a flexible and robust mechanism that could account for both narrow-band and broadband localisation on cones of confusion (Jin, 2001). Jin's model extends Humanski and Butler's approach to broadband stimuli by focussing on spectral contrast in neighbouring bands, and operates on frequency bands of varying bandwidth, emphasising the benefits of multi-scale analysis.

*Individualisation, detail, and ambiguity*

As every listener differs in terms of the shape of their head and ears, there has been much interest in the importance of individualisation of the spectral cues. In general it is agreed that a listener's auditory system is calibrated to his/her own ear-shape, during development and in adulthood if necessary (Hofman *et al.*, 1998), and that these learnt associations are responsible for accurate localisation. Accordingly, listeners make more localisation errors when presented with virtual stimuli based on another person's HRTFs (Wenzel *et al.*, 1993). Middlebrooks (1999a; 1999b) showed that these errors could be reduced by applying a simple frequency scaling to the non-

individualised HRTFs, essentially moving the prominent features of the HRTF to their appropriate position.

Whilst the unique frequency *positions* of HRTF features are crucial for accurate localisation, it has been shown that the auditory system is not necessarily dependent on fine spectral *detail* in the individualised HRTF. Kulkarni and Colburn (1998) asked subjects to distinguish real stimuli from virtual stimuli produced with systematic degrees of smoothing of the HRTF (done by reducing the number of Fourier coeffiecients for reconstruction). Only with very severe smoothing could subjects discriminate real stimuli from virtual stimuli, suggesting that fine spectral detail is not necessary. A similar conclusion was reached by Carlile and Leung (Carlile and Leung, 2001), who decomposed the magnitude response of measured HRTFs using principal components analysis (PCA), then recreated 'smooth' HRTFs using systematically fewer PCA coefficients. Two-dimensional localisation was found to be equally as accurate as control using only 20 out of 393 coefficients for reconstruction, which represents a substantial loss of detail.

Finally, a major ambiguity associated with the spectral cues requires addressing. For a listener to make use of the features of the HRTF, he or she must be able to distinguish them from spectral features of the source spectrum. This problem is often avoided in experimental situations by presenting stimuli with flat spectra, but natural stimuli are not always so simple. Indeed Rakerd *et al.* (1999) showed that subjects could not identify noises with different spectral shapes when they were roved on the MVP, presumably because they could not distinguish HRTF features from source features. However it has been shown that familiarity with a particular source spectrum can alleviate this confusion somewhat (Plenge, 1974). Furthermore, Macpherson and Middlebrooks (2003) reported that subjects can localise stimuli with a 'rippled' spectrum on the MVP under some circumstances. Certainly in natural situations the auditory system seems to have a solution to the problem of confounded source and directional spectra. Using the multi-scale frequency analysis model mentioned above (Jin, 2001), it was shown that good directional information is present even in very wide frequency bands. This broad-scale information is not as vulnerable to stimulus features as is a finer-scale analysis, and such a scheme may explain the robustness of the auditory system in locating natural stimuli.

**Neural coding of the spectral cues**

For spectral cues to contribute to a representation of auditory space in the nervous system, the system must be sensitive to location-dependent features such as peaks and notches. Poon and Brugge (1993) examined responses of auditory nerve fibres in the cat to broadband stimuli containing narrow-band notches at varying frequencies, and showed that most auditory nerve fibres were strongly inhibited by a notch at or near their best frequency. Most interestingly, some fibres were inhibited only by a notch *moving through* their best frequency and not a stationary notch. The authors proposed that this enhanced sensitivity might represent a mechanism for sound localisation during head motion.

In an attempt to inform the debate over whether peaks or notches are most important in human sound localisation, Moore *et al.* (1989) provided some data on how well listeners can detect changes in spectral shape in the high-frequency region. They found spectral peaks to be the more salient feature in terms of detection and discrimination, although changes in the centre frequency of both features were readily detected. They were in fact able to propose that discrimination of changes in the centre frequency of a prominent notch can explain spatial resolution of listeners in the frontal median plane.

The extraction of spectral cues for sound localisation by the nervous system has received relatively little attention in the literature (in comparison to the encoding of binaural cues). Spectral cues are complex, and it is difficult to manage the confound of source spectrum and location-dependent spectrum in a theoretical sense, let alone an experimental setting. However, the DCN has been implicated in this process as it appears to be specialised for spectral analysis.

Principal cells of the DCN are responsive to broadband stimuli and show sharp sensitivity to prominent features in the spectrum. Young and colleagues (Young *et al.*, 1992) demonstrated that type IV cells of the cat DCN respond to prominent location-dependent spectral notches produced by the pinna, and showed that this complex response profile is achieved though a combination of inhibition from type II neurons of the DCN and excitation from the auditory nerve (Nelken and Young, 1994; Nelken *et al.*, 1997). Disruption of the output fibres from the DCN causes deficits in elevation localisation in cats (May, 2000), confirming the contribution of this structure to spectral cue analysis.

Ascending DCN projections innervate type O units in the IC, and these neurons have been shown to shape DCN information (via local inhibition and excitation) to more selectively code for sound source location (Davis *et al.*, 2003). Sensitivity to spectral cues has also been observed in the primary target of the IC, the MGB (e.g. Imig *et al.*, 1997), as well as auditory cortex (e.g. Xu *et al.*, 1998).

### 1.3.4 Externalisation and distance perception

Another phenomenon that has been attributed to the spectral transformations provided by the auditory periphery is that of externalisation. Externalisation refers to the perception that a sound source is located outside of the head. Dichotic experiments using stimuli defined by an ITD and/or ILD (with no spectral cues) give rise to an 'internalised' impression, where the sound is perceived to be situated within the head. This experience should also be familiar to anyone who has listened to music over regular stereo headphones.

Plenge (1974) noted that externalisation can be achieved with headphones if an "ear-adequate" signal is presented. By this he meant a signal equivalent to one that might be received during true external presentation. Hartmann and Wittenberg (1996) also stressed that a plausible spectrum at each ear is required for externalisation. It is possible that dichotic stimuli fail to meet this criterion as ITDs and ILDs are simulated as being frequency-independent and, related to this, there is no spectral colouration. Externalisation can be achieved without fine spectral detail however (Kulkarni and Colburn, 1998). Most interestingly, externalisation can be achieved using non-individualised HRTFs (Wenzel *et al.*, 1993) while localisation is often severely disrupted.

Related to our ability to perceive a sound source as being located outside the head is our ability to judge how far from the head it is located. This chapter has focussed on how the auditory system can use the signals at the two ears to determine the *direction* from which an external sound is arising (usually defined by azimuth and elevation), and the perception of direction will also be the focus of the coming chapters. However, the dimension of *distance* is also extremely important for determining where a sound-emitting object is in relation to oneself.

In general, sound source distance is not judged very accurately by human listeners. Distance to near sources tends to be overestimated and distance to far sources underestimated (Zahorik, 2002). Several acoustic cues may be available depending on the source and the environment and these vary in their robustness. The intensity of a received sound is of course related to its distance, but this cue is confounded entirely with the presentation level of the sound. Nonetheless, overall level can give a crude idea of the distance of familiar sources and the relative distances of multiple sources (Mershon and King, 1975). There are also spectral cues that can indicate the distance of the source. For far sources, higher frequencies are attenuated significantly by the absorbing properties of the air (Blauert, 1983). For sources very close to the head, the spectral cues provided by the pinnae are no longer independent of distance and hence can provide cues to distance. Further to this, the binaural cues vary with distance for these very near sources (Brungart and Rabinowitz, 1999; Shinn-Cunningham *et al.*, 2000). In normal environments, echoes and reverberation can provide cues for sound source distance. The effect of reverberation on the 'spaciousness' in recorded music has long been noted, but more formal investigations have shown that the ratio of direct sound to reflected sound can indicate the distance of a sound source (von Békésy, 1960; Mershon and Bowers, 1979). It is likely that the auditory system combines and weights whatever distance cues are available in a flexible way; this weighting may depend on the stimulus and the environment (Zahorik, 2002) as well as the actual nature of the required task (see Zahorik and Wightman, 2001 for some discussion).

### 1.3.5 Space maps in the auditory system

It is clear from the discussions above that the representation of space in the auditory domain is a highly complex phenomenon. It becomes clear why there are so many levels of processing in the auditory system, as outlined in section 1.2.3. This simple overview has indicated that the acoustic cues to sound source location are processed in reasonably independent pathways, via binaural interaction and spectral analysis, at least in the early processing stages. However, psychophysical experiments have shown that each of the cues is important for listeners to accurately localise a sound source (see section 1.4). It seems that the cues must be combined at some stage in the

auditory system to allow a coherent estimate of location. Perhaps the question can be phrased: how *complete* is the neural representation of space? Many researchers have been interested in finding a "map" of auditory space, where the cues defining sound source location are combined into an explicit representation. Such a map is defined by an orderly relationship between neural units and their preferred directions in auditory space.

The clearest demonstration of a spatiotopic map in the auditory system is in the deep layers of the SC, where sound source azimuth is represented across the rostrocaudal axis of the nucleus and elevation is represented across the mediolateral axis (ferret: King and Hutchings, 1987; guinea pig: Palmer and King, 1982; cat: Middlebrooks and Knudsen, 1984). This map is superimposed on analogous maps of visual space and is involved in cross-modal interactions and the driving of motor responses to novel stimuli.

To achieve the spatial selectivity seen in SC neurons, binaural and monaural information must be integrated and thus these neurons are broadly tuned to frequency with no apparent tonotopicity (Wise and Irvine, 1983; Carlile and Pettigrew, 1987). Furthermore, the spatiotopic map appears to arise from a systematic variation across the nucleus of binaural and monaural cue sensitivities (Wise and Irvine, 1985; Carlile and King, 1994; Palmer and King, 1985). So it appears that in the SC the acoustic cues to sound source location are integrated to form a physical neural map of auditory space. It is possible, however, that this orderly representation of space is necessary to enable interactions with the extremely orderly visual and motor space maps in the SC, and does not represent an optimal way of encoding auditory space in the brain. Indeed, searches for orderly representations of auditory space in primary auditory nuclei have been largely unsuccessful. In the IC, many neurons are selective for a restricted region of space (e.g. Delgutte *et al.*, 1999) but these neurons do not appear to be organised topographically (see Sterbing *et al.*, 2003; Oliver *et al.*, 2003). There is some evidence for spatiotopicity in the external nucleus of the IC (Binns *et al.*, 1992) and the brachium of the IC (Schnupp and King, 1997; King *et al.*, 1998), which project to the SC, but these are less specific than in the SC and are probably involved in the derivation of the accurate SC map.

In the cortex, there is no evidence for spatiotopic maps, and most neurons respond to a large range of locations in space. Investigations in cats indicate that non-spatiotopic mapping strategies may exist in this higher centre, as spatial information is

available in onset firing latencies (Brugge *et al.*, 1996) and temporal firing patterns across ensembles of cortical neurons (Middlebrooks *et al.*, 1994; Middlebrooks *et al.*, 1998).

Some insight into the organisation of spatial information in human auditory cortex has come from non-invasive neuromagnetic studies. In a recent example, measurements of auditory-evoked magnetic fields on the scalp of human listeners showed that different cortical areas mediate localisation in azimuth (primarily based on binaural cues) and elevation (primarily based on monaural spectral cues). While azimuth appears to be analysed in the contralateral hemisphere, the right auditory cortex appears to dominate in the analysis of elevation. Furthermore, the latency of responses is greater for the processing of elevation (Fujiki *et al.*, 2002). These differences suggest that the cues to sound source location are not integrated into a simple map of two-dimensional space in auditory cortex, but rather that distinct processing schemes exist. Nonetheless, it is clear that the different cues to sound source location (as well as information about the nature, identity, loudness and proximity of the source) are 'bound' in some way to give rise to a coherent perceptual object.

## 1.4 Human spatial performance

There is an enormous body of literature, spanning over a century now, that is concerned with how accurately human listeners localise sounds (there is also a range of papers describing localisation behaviour in other animals e.g. cat: Casseday and Neff, 1975; monkey: Brown *et al.*, 1982; barn owl: Poganiatz *et al.*, 2001). This section will briefly review a small portion of the human literature, describing both absolute and relative localisation studies. The described studies all examined directional localisation, and distance perception will not be covered here. These observations on human performance are largely interpretable in light of the previous sections on the cues for sound source location. As discussed, binaural cues enable localisation in the horizontal dimension, and spectral cues enable resolution of the cones of confusion. In general then, performance in these dimensions is related to information carried by the respective cue or cue sets. As the spectral cues are a particularly vulnerable cue, cone of confusion errors are commonly observed; these

include elevation errors as well as the common front-to-back confusion (Makous and Middlebrooks, 1990; Carlile *et al.*, 1997).

### 1.4.1 Absolute localisation

**Localisation of broadband stimuli**

The optimum stimulus for the auditory localisation system appears to be a single broadband sound having a relatively flat spectrum presented in anechoic space. Under these conditions, errors are small and confusions are rare. This has been demonstrated in many studies, including Carlile *et al*. (1997) using a nose-pointing response method and King and Oldfield using a 'shotgun' pointing method (King and Oldfield, 1997). Errors are generally on the order of a few degrees, and are smaller in the anterior portion of space, especially near the audiovisual horizon (Carlile *et al.*, 1997).

For broadband stimuli with non-flat spectra (particularly when the spectrum is unfamiliar), errors are more common as a result of impaired spectral cue extraction. When the spectrum of a noise stimulus is 'scrambled' by introducing amplitude changes on a critical band scale, spectral cue localisation is robust to changes of ± 5dB (Kulkarni and Colburn, 1998) but not ± 40 dB (Wightman and Kistler, 1997). Scrambling on the order of ± 20 dB has been reported to have little effect on localisation (Wightman and Kistler, 1989a; 1992) although unpublished data from our laboratory shows that this level of spectral scrambling results in a 3-fold increase in cone of confusion errors. Ripple spectrum stimuli with peak-to-trough depths of 40 dB also have an adverse effect on vertical localisation, but only on certain frequency scales (0.5 – 2 ripples per octave), suggesting that the auditory system is in fact quite robust to a range of source spectrum features (Macpherson and Middlebrooks, 2003).

There are some minimum requirements for the accurate localisation of broadband sound sources. Hofman and van Opstal (1998) showed that at least 80 ms of broadband noise is required before elevation localisation is robust, although horizontal (binaural) localisation is stable after only 3 ms. It was found that the noise did not have to be sustained however. Repeated trains of 3ms noise bursts could give good elevation performance but only for short inter-burst intervals (less than 10 ms) suggesting that short-term estimates are integrated in a leaky fashion.

An important broadband signal of particular relevance to human listeners is speech. The localisation of this spectrally and temporally complex signal is dealt with at some length in Chapter 5.

**Localisation of bandlimited stimuli and pure tones**

A lot of the early work on human sound localisation was carried out using pure tones (Stevens and Newman, 1936). In general, pure tones are localised reasonably reliably in the lateral dimension due to the robustness of the binaural cues. As discussed earlier however, the ITD and the ILD are most useful in different frequency ranges and hence performance depends on the frequency of the tone (Sandel *et al.*, 1955). Importantly, it is impossible to resolve the cone of confusion for a pure tone without head movements, as there is no scope for across-frequency spectral analysis. Thus front-back ambiguities are common and elevation perception is poor as discussed previously.

Low-pass filtered noise is localised well in the horizontal dimension, but extremely poorly on cones of confusion due to the absence of high-frequency pinna cues. Severe disruptions of localisation have been reported for upper cut-offs of 2 kHz (Carlile and Delaney, 1999) and 3 kHz (Butler and Humanski, 1992). King and Oldfield (1997) progressively low-passed white noise and measured localisation performance for a range of locations in 2-dimensional anechoic space. They reported that an upper cut-off of 9 kHz was needed for accurate elevation localisation but that frequencies above 10-13 kHz were required for reliable front-back discrimination. A critical observation was made by Perrett and Noble (1995) concerning the collection of localisation responses in situations where cone of confusion errors are to be expected. Their general argument was that the range of available response options can affect responses and hence performance measures. They demonstrated this idea explicitly by presenting three types of stimuli (broadband pink noise, 400 Hz low-pass filtered noise, and 400 Hz pure tones) from a lateral plane under two conditions. In the first, a single vertical speaker array was used. All three types of stimuli were localised well on this array. In the second, a vertical and a horizontal speaker array were used. In this case, broadband stimuli were localised well but the low-frequency stimuli were often localised to the incorrect array, representing typical cone of confusion errors. This showed that in the first condition (which replicated one of the experiments by

Butler and Humanski, 1992), subjects were able to resolve such errors on the basis of the available responses.

High-pass filtered stimuli are localised more accurately than low-passed stimuli in general. Good performance in all dimensions has been reported for lower cut-offs of 2 kHz (Carlile and Delaney, 1999) and even up to 6-9 kHz (King and Oldfield, 1997).

**Monaural localisation**

It was suggested in 1901 that listeners could localise reasonably in the horizontal plane using only one ear (Angell and Fite, 1901) and the authors suggested that high frequency spectral cues were used. However, subjects in this study required training and the resulting errors were still large (18° error). A later study Butler *et al.* (1990) demonstrated that monaural listeners could *not* localise in the horizontal plane, but that elevation estimates were near-normal on the side of the functioning ear.

However, Wightman and Kistler (1997) made several good criticisms of previous monaural localisation studies. They carried out their own free-field monaural experiment and compared the results to a single-ear virtual stimulus experiment. They found that monaural results were reasonable in the free-field experiment but localisation ability was completely abolished in the virtual condition, suggesting "inadequate monauralization by plug and muff". Their claim that previous studies have been tainted by this inadequacy is consistent with all but one study, where unilaterally deaf listeners were able to localise well even on the side of the deaf ear (Slattery and Middlebrooks, 1994). These listeners were no doubt highly adapted to monaural listening and bring to mind Angell and Fite's listeners (1901), who reportedly could only perform the monaural task after training.

**Localisation with echoes and reverberation**

In everyday environments, walls and objects give rise to echoes and reverberation that clearly disturb the acoustical cues to location arising at the two ears (see Shinn-Cunningham and Kawakyu, 2003). Surprisingly, localisation is fairly robust under these conditions (Hartmann, 1983; Shinn-Cunningham, 2000). It has long been thought that a phenomenon known as the 'precedence effect' is involved in the suppression of echoes, and perhaps could play a role in maintaining adequate localisation in echoic environments (for reviews see Zurek, 1987; Litovsky *et al.*,

1999). However ongoing modelling work by Shinn-Cunningham and colleagues is suggesting that integration over time of noisy location estimates (weighted by certainty) can lead to reasonable final estimates of sound source location (Shinn-Cunningham and Kawakyu, 2003).

### 1.4.2 Relative localisation

In 1958, Mills defined the 'minimum audible angle' (MAA) as the smallest detectable change in sound source position for which a listener can identify the direction of change correctly on 75% of trials (Mills, 1958). This was measured by presenting a pair of sequential stimuli and asking subjects to determine the direction of displacement (left or right, in the horizontal plane). For pure tones, this measure was found to vary greatly with frequency and location, but reached about 1° at best.

Since the early work of Mills, many researchers have measured the MAA under various stimulus conditions. The MAA is generally improved with the use of wideband sources rather than tones (e.g. Recanzone *et al.*, 1998). Furthermore, the MAA using low-passed noise (70 Hz – 2 kHz) or high-passed noise (6 - 15 kHz) has been shown to be equivalent to that using broadband noise (70 Hz – 15 kHz) in the horizontal plane. Thus it seems that with a sufficiently wide bandwidth, either ITD or ILD cues can produce optimal horizontal resolution.

There is some debate as to whether the horizontal MAA varies as a function of azimuth, as absolute localisation is known to do. Mills reported a severe degradation of performance using tones for azimuths of 30°, 60° and 70° as compared to that measured at 0° (Mills, 1958, 1972). If one assumes that the relative localisation of pure tones is dependent on a change in ITD (low frequencies) or ILD (high frequencies) then this finding makes sense as the binaural rate-of-change is lower at lateral locations. However, Harris (1972) compared the MAA for tones at 30° and 60° azimuth to that at 0° and found no difference.

For broadband stimuli, it is reasonable to expect that the dependence on binaural changes is reduced due to the availability of monaural spectral cues to detect location changes. Saberi *et al.* (1991b) used broadband noise and noted a small increase in the MAA from 1° at the frontal location to 5° at a location directly to the side. In a more recent study, the MAA for noise was measured at several azimuths

between 0° and 48°, and there was no evidence of variation in this range (Recanzone *et al.*, 1998). A criticism of that study, however, is that the loudspeakers employed were placed at 4° intervals. This seriously limits the resolution with which the MAA can be examined.

The involvement of the spectral cues is particularly important when considering the MAA in the vertical dimension. Although few studies are available on this topic, the general finding is that the vertical MAA is larger than the horizontal MAA when measured in the front (Perrott and Saberi, 1990; Strybel and Fujimoto, 2000) and does not change considerably with lateral position (Saberi *et al.*, 1991b). In a recent report, absolute and relative localisation on the median vertical plane was examined in blind listeners (Lewald, 2002). These listeners showed errors in absolute localisation, presumably because a lack of visual feedback has prevented optimal association of the monaural spectral cues with appropriate locations in space. Interestingly, however, these blind listeners produced MAAs on the median vertical plane that were as good as sighted listeners. This emphasises that relative spatial acuity is highly dependent on the detection of *changes* in the acoustic cues to location rather than on the detection of location itself.

## 1.5 Aims and overview

The previous review demonstrates the wealth of research that has built our current knowledge on human sound localisation. Much is now understood about the physical cues to sound source location and the neural mechanisms that encode these cues, and studies of human spatial performance have been crucial in developing this understanding. In parallel with work on sound localisation, a huge and extraordinarily dense body of literature concerned with multiple-source listening has evolved. The focus has primarily been on complex (and important) problems such as the masking of one sound by another, speech intelligibility in noise, and auditory scene analysis. Some of this literature will be discussed throughout the thesis. However, from a spatial hearing perspective, there appears to be a gap or a missing link between these two bodies of research. Very few studies have attempted to systematically examine spatial performance with multiple auditory objects.

It is intriguing to discover whether what we know about sound localisation holds up in multiple-source environments. As competing sounds must share a common sensory apparatus, and presumably common processing pathways, interactions in their spatial representation must be expected. Indeed Bregman in his book on auditory scene analysis (Bregman, 1990, p. 312) proposed that cues associated with segregation ("world structure cues") should be named as another localisation cue.

The overall goal of the work described in this thesis was to systematically investigate spatial hearing with concurrent sound sources. The experiments were concerned with how accurate human listeners are at spatial tasks involving more than one source. The general approach was to use sources that overlapped in *both frequency and time*. This is a realistic situation, but also one that maximally challenges the auditory system to reveal perhaps its most sophisticated abilities. Another key component of the approach was to examine performance in both horizontal and vertical spatial dimensions, in order to involve the different localisation cues. Most notably, the role of spectral cues provided by the pinnae in multiple source listening is poorly understood.

The following chapters describe and discuss a series of experiments addressing these issues. Chapter 2 introduces the experimental environment and in particular describes the generation of individualised virtual auditory space. Chapters 3 to 7 describe the psychoacoustic experiments, covering the relevant literature as well as the details and findings of the work. Each of the experimental chapters is concerned with a specific aim or set of aims as described below.

In Chapter 3 the aim was to measure auditory spatial resolution. Using a two-point discrimination approach borrowed from the visual and somatosensory psychophysical literature, the ability of listeners to resolve two simultaneous broadband point sources was examined. Resolution was examined for different spatial configurations and in different regions of space.

In Chapter 4 the aim was to directly examine how the spatial percept of an object is influenced by the presence of a concurrent object. Localisation of a broadband sound was measured in the presence of another broadband sound, with the two distinguishable on the basis of their temporal characteristics. The influence of several parameters was examined, such as simultaneity of the pair, stimulus duration, relative characteristics of the two stimuli, and the spatial configuration of the pair.

In Chapters 5-7 the focus was turned to speech, a complex natural sound. It was first necessary to examine how well speech is localised in isolation, to provide a baseline for the subsequent examination of simultaneous situations. Thus in Chapter 5, localisation accuracy with single monosyllabic words was examined. Of particular interest was the importance of the high-frequency information that is present in speech.

In Chapter 6, the approaches used in Chapters 3 and 4 were adopted in order to examine spatial performance with concurrently spoken words. The aim was to reveal how the spatial auditory system deals with competing natural signals that vary independently in their temporal and spectral structure.

In Chapter 7 the aim was to briefly explore how spatial resolution for concurrent sources might be relevant in a functional setting, where concurrent streams of speech need to be segregated. The effect of spatial separation of competing talkers on the ability of listeners to attend to a target talker was examined. The role of the different spatial cues, particularly the spectral cues, in this kind of task was explored.

Finally, in order to tie together the findings of the different experimental chapters, Chapter 8 provides a summary and a general discussion. Here the results as a whole are discussed in relation to auditory spatial processing and theories of auditory object analysis.

# Chapter 2: The experimental environment

## 2.1 Subjects

In total, 10 subjects participated in the experiments comprising this thesis: 6 females and 4 males aged between 24 and 36 years. Audiometric thresholds were measured using a clinical audiometer, and all subjects had normal hearing in both ears across the standard testing range (250 - 8000 Hz). The subjects are uniquely identified by a number (S1 – S10) and are referred to as such throughout the thesis. In each experiment, only a subset of the subjects was employed, and they are specified in the appropriate sections. Subjects S1 - S5 had participated in previous auditory psychophysical experiments in the laboratory, whereas subjects S6 - S10 were relatively inexperienced. Each subject was given a thorough description of what was expected of him or her prior to experimentation and all procedures were in compliance with the University of Sydney Human Ethics Committee guidelines.

All subjects underwent a standard period of localisation training and testing (see sections 2.4.1 and 2.4.2) before participating in the experiments comprising this thesis. In addition, each subject underwent a recording session during which their outer ear directional transfer functions were measured (section 2.5.2) for the purposes of creating an individualised virtual auditory environment.

## 2.2 The spatial co-ordinate system

Source positions used for stimulus presentation throughout this work were all located on a sphere of space with radius 1m. In each series of experiments, locations on this sphere are described using one of two co-ordinate systems.

In the first description (Figure 2.1a), a single pole is oriented vertically through the centre of the sphere, and positions are defined by their *azimuth* and *elevation*. Azimuth describes horizontal angular displacement due to rotation about the pole: 0° azimuth is located on the frontal midline, 90° and -90° azimuth are to the

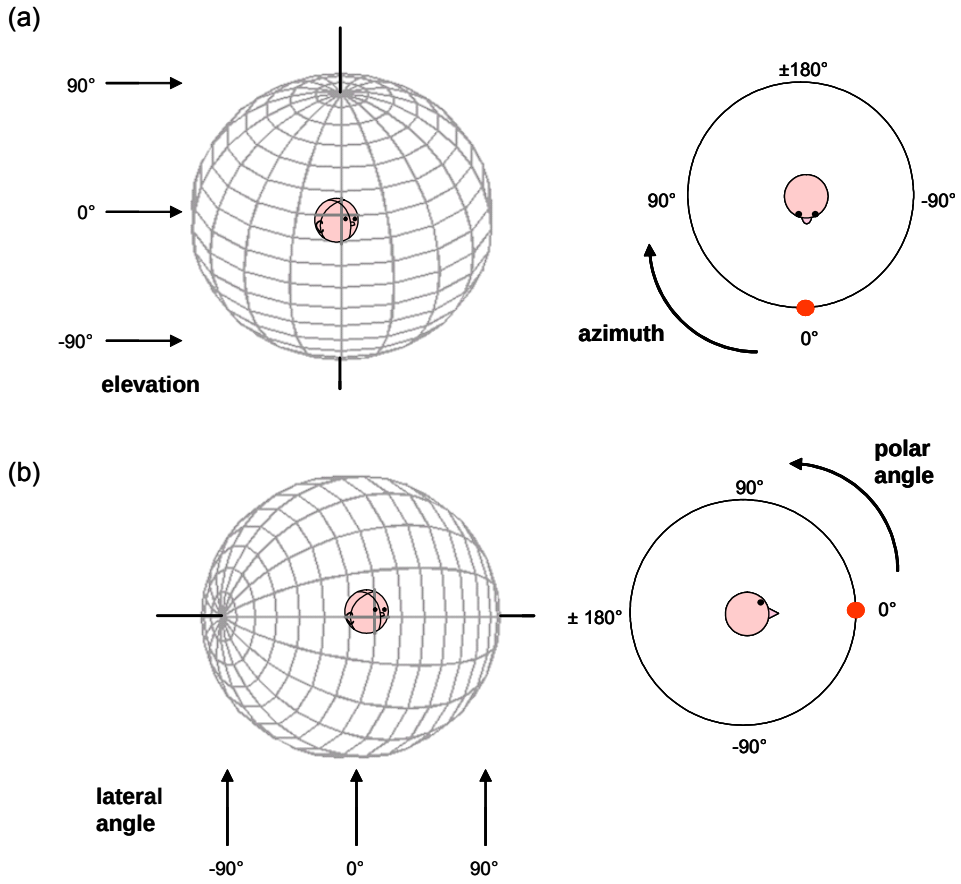



**Figure 2.1** (a) The azimuth/elevation co-ordinate system consists of a single pole oriented vertically. Azimuth describes horizontal angular displacement due to rotation about the pole (right illustration) with 0° azimuth at the front and 180° and -180° azimuth coinciding at the rear. Elevation describes horizontal planes that pass through the single pole (left illustration) with 0° elevation passing through the two ears and 90° and -90° elevation defining the upper and lower limits of the sphere. (b) The lateral/polar co-ordinate system consists of a single pole passing through the two ears. The lateral angle is the horizontal angle away from the midline (left illustration), with 0º defining the median sagittal plane and -90º and 90º defining the left- and right-most extremities of the sphere. The polar angle describes the angle around the circle described by a particular lateral angle (right illustration). A polar angle of 0º describes the front-most location on this circle, with polar angle increasing to 90º at the top and -90º at the bottom, with 180º and -180º coinciding at the back.

right and left respectively on the interaural axis, and 180° and -180° azimuth coincide on the rear midline. Elevation describes horizontal planes that pass through the single pole: 0° elevation describes a plane passing through the two ears, 90° elevation is the top of the sphere, and -90° is the lowest elevation. This co-ordinate system is an intuitive one when considering a human listener orienting themselves to face a sound

played from around them (a task used in the localisation paradigm, see section 2.4.1). A natural behaviour would be to rotate the body horizontally (i.e. change azimuth) and to nod the head up or down (i.e. change elevation).

In a second description (Figure 2.1b), the pole is oriented horizontally through the two ears, and positions are defined by their *lateral* and *polar* angles. The lateral angle is the horizontal angle away from the midline, where negative lateral angles (down to -90º) correspond to the left hemisphere of space and positive lateral angles (up to 90º) correspond to the right hemisphere of space. In this system the front-back dimension and the elevation dimension are encompassed in the polar angle, which is the angle on the circle described by the lateral angle. A polar angle of 0º describes the front-most location on this circle, with polar angle increasing to 90º at the top and 180º at the rear aspect. In the lower hemisphere, polar angle decreases to -90º at the lowest point of the circle and -180º coincides with 180º at the back. The lateral-polar co-ordinate system is a particularly convenient framework for describing auditory localisation data because each lateral angle corresponds approximately to a particular binaural interval, and the set of polar angles describe locations on a single cone of confusion.

## 2.3 Testing facilities

### 2.3.1 Anechoic chamber

Recording of directional transfer functions and localisation testing were carried out in an anechoic chamber (Figure 2.2a) of 4m x 4m x 4m (effective working area 2.5m x 2.5m x 2.5m). The chamber had an insertion loss of better than 30dB for sound frequencies greater than 100Hz rising rapidly to greater than 60dB above 500Hz. Sound absorbing wedges lining the walls, floor and ceiling ensured that 99% of incident sound energy for frequencies above 200 Hz was absorbed. Within the chamber, a robotic hoop system carrying a stimulus loudspeaker (VIFA-D26TG-35) could be moved to position the loudspeaker at any location on an imaginary sphere of radius 1m. Two semicircular hoops comprised this system: the outer hoop was suspended from the ceiling of the chamber and defined the azimuthal placement of the loudspeaker, and the inner hoop was attached to the outer hoop and moved to position

(a) (b)

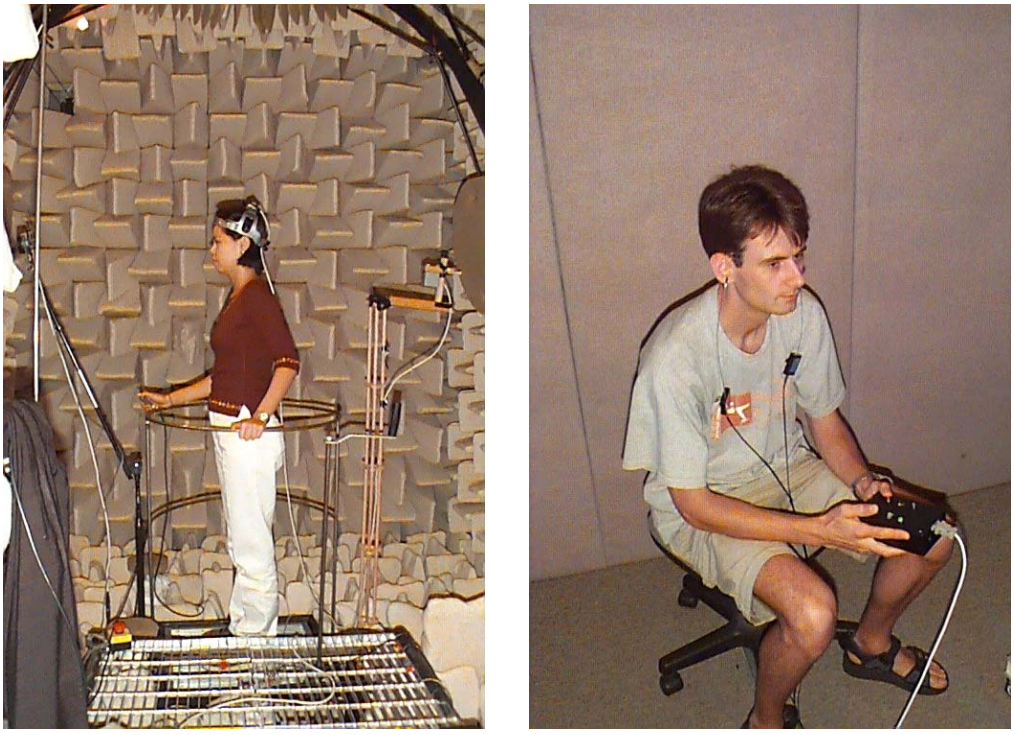


**Figure 2.2** (a) The anechoic chamber. The walls/floor/ceiling of the chamber are lined with sound absorbing wedges and the hoop system can be moved to place the loudspeaker at any location on a 1m radius sphere. A subject is shown standing on the metal platform and wearing the electromagnetic head-tracker. (b) The soundproof room. Sound attenuating material lining the walls and ceiling excludes external noise. A subject is shown seated in the chamber, wearing the in-ear tubephones and holding the response box.

the loudspeaker at any elevation (excluding elevations greater than 50° below the centre of the sphere). Placement of the loudspeaker was automated and accurate to within 1°.

The hardware within the chamber was interfaced to a TDT System II rack (Tucker-Davis technologies) which communicated with a PC. Sound stimuli were generated offline and passed through a digital-to-analogue converter (80 kHz, TDT: DD1) and a programmable attenuator (TDT: PA4) before delivery. Stimuli were presented either via the hoop loudspeaker or via earphones. Stimulus generation and data collection were controlled via the PC using MATLAB software (Mathworks Inc., versions 5.5 to 6.5).

A metal platform to support subjects (seated or standing) was located at the base of the sphere defined by the hoop system. The platform was supported above the

floor of the chamber and its height could be altered by means of a winch in order to position a subject's head in the centre of the sphere. The chamber was fitted with a two-way intercom to allow communication between the subject and the experimenter when required.

### 2.3.2 Soundproof Room

Experiments that did not require an anechoic environment or the hoop system were carried out in a small soundproof room. This room could be sealed to eliminate all light, and sound attenuating material lining the walls and ceiling offered approximately 40dB attenuation of external noise. This chamber was fitted externally with a PC and TDT System II rack identical to those associated with the anechoic chamber, and stimuli were delivered via earphones to the subject who was seated on a chair in the chamber (Figure 2.2b). A laptop computer or a hand-held response box was provided as required for the various experiments. The response box had two buttons, each coupled to a light emitting diode (LED) which lit up to confirm a response.

## 2.4 Localisation paradigm

### 2.4.1 Basic testing procedure

As the experiments described in this thesis were largely interested in examining spatial perception in human listeners, it was important to be able to measure spatial capabilities in an accurate and objective way. To this end, a standard sound localisation testing paradigm was used in which subjects localised auditory stimuli presented in darkness using a head-pointing technique (Carlile *et al.*, 1997).

For localisation testing, subjects were required to stand comfortably on the platform in the anechoic chamber (Figure 2.2a), and a circular metal handrail was mounted at waist level to (a) constrain the subject to the centre of the hoop system, and (b) act as a guide for the subjects to orient themselves during standing and turning on the platform. To monitor head position throughout localisation testing, an electromagnetic tracking device was used (Polhemus 3SPACE ISOTRACK II). This

device consisted of a transmitter, mounted on a wooden frame behind the subject, and a receiver fitted to a plastic frame worn on the subject's head. Before a test, the subject was positioned in the centre of the chamber and aligned with the hoop co-ordinate system to ensure that the orientation of their head could be meaningfully expressed in terms of stimulus positions on the hoop. This was done by calibrating the head-tracker when the subject was positioned in the centre of the chamber with their nose pointing straight ahead to the 'zero' position (0° azimuth, 0° elevation). A pair of lasers mounted on the inner hoop aided this alignment: one mounted at the zero position could be fixed on the subject's nose to align them laterally, and the other mounted directly to the left (at -90° azimuth, 0° elevation) could be fixed on the subject's ear canal to align them in the front-back dimension. A head position indicator was located in front of the subject to provide feedback as to the position and orientation of the head. This took the form of a series of coloured LEDs which were linked to the head-tracking system to give a visual reference to head position. A central green light lit up when the subject's head was accurately centred and pointing to the 'zero' position. Four red lights located around this central one lit up to indicate deviations in azimuth and elevation from this position.

To commence a localisation session, the subject aligned their head with the hoop co-ordinate system using the LED display and pressed a hand-held response button when ready. Importantly, this ensured that the subject's gaze was directed straight ahead at the commencement of the test; it has been shown that eccentric gazes can bias sound localisation judgements (Lewald, 1998; Getzmann, 2002). A stimulus was then delivered from the loudspeaker, which was located at a random position on the sphere defined by the hoop system, and the subject was required to indicate his/her perceived location by pointing with the nose and pressing the response button. The location indicated by the subject was automatically obtained from the head-tracking system at the time of the button press. The LED feedback display was then activated again and the subject was required to align his/her head again for the next trial. In any test this process was repeated for a predetermined selection of stimulus locations.

### 2.4.2 Localisation training

Before localisation ability could be reliably measured, subjects had to undergo a training protocol to become familiar with the localisation task. In particular, this includes the head-pointing response technique, which requires that the subject point his or her nose toward the perceived location of a stimulus. Training was designed to eliminate the tendency to 'capture' a sound with the eyes rather than point the nose to the target (Carlile *et al.*, 1997). Stimuli used for training were 150ms broadband Gaussian white noise bursts (with 10ms raised-cosine onset and offset ramps) presented from the loudspeaker on the hoop. This duration was brief enough that subjects could not make head movements during the stimulus presentation, ensuring that the task was a static not a dynamic one (dynamic cues can improve localisation, see Wightman and Kistler, 1999).

A training test followed the basic procedure outlined in the previous section, with the addition of two feedback steps after each localisation response. The first was visual feedback, provided in the form of a small LED located on the loudspeaker, which the subject was allowed to use to readjust the head if needed and respond again. This was then followed by auditory feedback, where the sound was presented repetitively allowing the subject to further reorient their head. The accuracy of the orientation of the head was indicated by the frequency of pulsation of the stimulus. When the rate was maximal, this informed the subject that they were localising accurately and the response button was pressed. In any one training block there were 36 different stimulus locations presented. After a training test, localisation performance was examined using the analyses described in the following section. Training was continued until performance defined by these measures had reached a plateau, indicating that optimum performance had been reached.

### 2.4.3 Analysis of localisation performance

Upon completion of any localisation test, the pattern of responses was visually inspected using a custom-written MATLAB function. This function generated a graphical user interface with a rotatable sphere representing the sphere defined by the hoop system. Subject response locations were plotted in relation to target stimulus

locations on this sphere, enabling the nature and distribution of errors to be examined interactively.

In order to grade the overall level of accuracy of a set of localisation data, two objective measures were used. The first of these was the spherical correlation coefficient (SCC, see Fisher *et al.*, 1987) and its use with localisation data is described in detail elsewhere (Leong and Carlile, 1998). Briefly, it describes the degree of correlation between target and response locations based on their directional cosines (1 = perfect correlation; 0 = no correlation). This measure is useful in quantifying the 'global' accuracy of a set of localisation responses.

It is not uncommon for localisation data to contain a small subset of large, specific errors associated with cones of confusion (refer to section 1.4). For example, if the monaural spectral cues do not sufficiently resolve the ambiguity associated with a binaural interval (for whatever reason) a subject will commonly make 'front-back confusions' (confuse the true location with a location mirrored across the interaural axis). This can manifest in the data as a bimodal distribution or a consistent mislocalisation, either of which may degrade the SCC in a misleading way. For this reason, target and response locations were analysed in terms of the lateral-polar angle co-ordinate system (section 2.2), and cone of confusion (COC) errors were defined as trials in which the target and response polar angle differed by more than 90°. These trials were removed from the set for the calculation of the SCC, and the occurrence of this type of error (expressed as a percentage of the total number of trials) was used as a second measure of accuracy to complement the SCC.

## 2.5 Individualised virtual auditory space

### 2.5.1 The use of virtual auditory space

There are several techniques for presenting spatial auditory stimuli to listeners, and many examples used in research were given in Chapter 1. The primary technique is to present sounds from external loudspeakers to imitate a real listening environment ("free-field listening"). However more recently there has been a surge in the use of "virtual listening", where a realistic spatial percept akin to the free-field experience is achieved using headphone presentation. This approach goes by several names but is

referred to here as virtual auditory space (VAS, for review see Carlile, 1996b). All of the experiments comprising this thesis utilised a VAS environment.

The basic approach in this technology is to accurately simulate the wave pattern at each eardrum occurring after stimulation with an external sound source. This involves presenting the sound of interest over headphones after simulating the natural distortions of the signal that would be produced by the environment (echoes, reverberation) and the auditory periphery of the listener (filtering by the head, shoulders, and outer ears). Importantly, these cues vary as a function of the relative locations of source and listener, and do so independently for the left and right ear. Furthermore, as the anatomy of an individual's auditory periphery is unique, the transfer function from source to ear canal is individualised. To generate VAS for a particular individual then, it is necessary to measure the transfer functions for their ears for every source location that is required. Such measurements are done routinely in a number of laboratories using a variety of techniques (see for example Wightman and Kistler, 1989b; Pralong and Carlile, 1994; Middlebrooks *et al.*, 1989; Møller *et al.*, 1995) and the techniques relevant to the work in this thesis are described in the sections to follow.

VAS is an extremely useful tool for basic research into spatial hearing. It allows the fast and flexible delivery of stimuli, where the placement of objects in auditory space can be done accurately and with absolute repeatability. For the presentation of concurrent sounds in various configurations (as described in Chapters 3, 4, 6, 7) this was an essential tool, as it avoided the need for multiple speakers and an extremely sensitive placement system. In addition, easy and independent manipulation of the stimuli reaching the two ear canals provides unique control over the spatial cues received by the auditory system. This gives one the opportunity, for example, to isolate a specific spatial cue and examine its role in a particular auditory perceptual process - a capability that does not exist in a traditional loudspeaker set-up.

### 2.5.2 Measurement of directional transfer functions

To render individualised VAS for each subject, directional transfer functions (DTFs) were measured. The DTF describes the directional filtering properties of the auditory periphery, and varies with the direction of an incoming sound relative to the head.

DTFs were obtained by recording impulse responses, transposing these to the frequency domain using a fast Fourier transform, and removing location-independent components (associated with the microphones and the recording environment, see below). To obtain a good representation of space, impulse responses (and DTFs) were obtained for 393 positions spaced approximately evenly on the sphere defined by the hoop system. Recording positions were distributed between 15 different planes of elevation (from 90° to -40° in 10° steps as well as -45°). Azimuthal spacing of positions in each of these planes ranged from 10° to 45° depending on the size of the azimuthal circle.

A blocked ear canal technique was employed to record impulse responses, described in detail elsewhere (Møller *et al.*, 1995). Briefly, a small microphone (Sennheiser KE-4-211-2) was placed inside each ear canal such that its face was flush with the distal end of the ear canal (Figure 2.3a). The microphone was wrapped in soft surgical tape to give a comfortable fit within the canal, and the wire from the microphone was fed unobtrusively out of the ear and taped to the neck of the subject. The microphones, once inserted, were fed to custom-made pre-amplifiers that were strapped to the torso of the subject.

Subjects were seated in the centre of the anechoic chamber and aligned to the stimulus co-ordinate system as described previously (section 2.4.1). For the duration of the recording period (approximately 40 minutes) subjects were required to keep their head still, and to assist with this they were seated on a chair fitted with a moulded chin rest (Figure 2.3b). A hand-held button allowed the subject to stop and re-commence the recording session at any time if they needed to move or rest.

The recording stimulus was a 1024 bit Golay code pair presented 12 times, with the resultant input averaged to increase the signal-to-noise ratio (Golay, 1961; Zhou *et al.*, 1992). To minimise artifacts that may arise due to small head movements during the Golay code pair repetitions (Zahorik, 2000), head position was continuously monitored using the head-tracker and recording stopped immediately when deviations greater than 2° from the calibrated position occurred. Subjects were instructed on how to re-adjust their head using the visual feedback display (section 2.4.1). The output of the recording microphones was subjected to anti-alias filtering at 30 kHz (TDT:FT6) and analogue-to-digital conversion at 80 kHz (TDT:DD1).

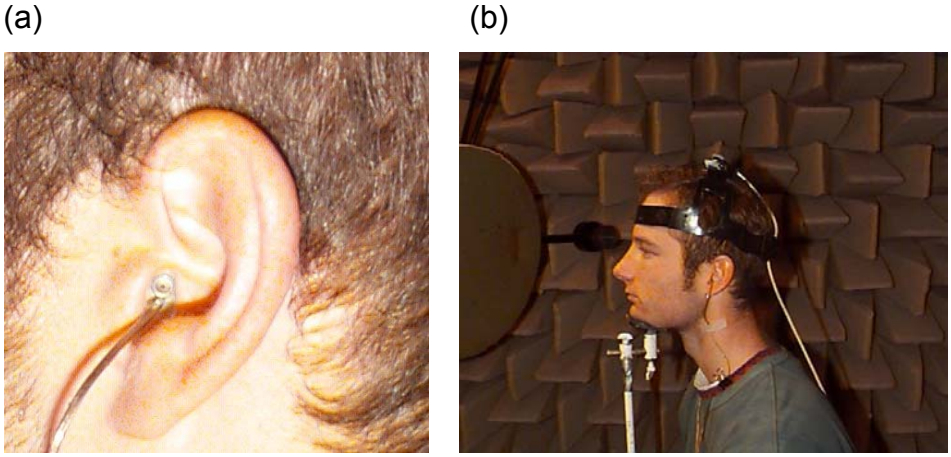



**Figure 2.3** Recording of directional transfer functions. (a) The blocked ear canal technique involves the placement of a small microphone inside the ear canal such that its face is flush with the entrance to the ear canal. The wire from the microphone feeds unobtrusively out of the ear. (b) For the recording session, subjects were seated in the centre of the anechoic chamber with microphones in each ear canal. Their head position was kept steady with the aid of the chin-rest and feedback from the head-mounted head-tracker (both shown).

Immediately before or after a recording session, a calibration recording was carried out to characterise the microphone and system transfer functions. This procedure was identical to the DTF recording procedure, but the microphones were positioned alone in the centre of the chamber. These recordings were deconvolved from the recordings made in the subject's ear canals to leave only transformations attributable to the auditory periphery (Pralong and Carlile, 1994). Finally, location-independent components (components that are constant across locations) were removed from the recorded transfer functions (Middlebrooks and Green, 1990). This was done by dividing the ear canal recordings by the mean magnitude spectrum across all recording locations to leave only the directionally-dependent DTFs. The DTFs were used to generate finite impulse response filters with a minimum phase response, on the basis of several reports that the spatial impression is not dependent on frequency-dependent phase differences (Kistler and Wightman, 1992; Hartmann and Wittenberg, 1996; Kulkarni and Colburn, 1998). ITD was calculated from the left and right ear impulse response recordings using the peak of the cross-correlation function and was applied to the DTF filters by delaying the lagging ear appropriately. The filters were then band-passed from 300 Hz to 16 kHz. A 'virtual' stimulus at a

particular location could then be created for an individual listener by convolving the desired signal with the appropriate left and right ear filters before delivery.

### 2.5.3 Stimulus delivery

Virtual auditory space stimuli were presented using in-ear headphones (Etymotic Research ER-2) with no headphone compensation. The frequency response of each headphone is flat up to 10 kHz according to manufacturer's specifications, and it has been reported that their response varies by less than ±5dB between 500 Hz and 10 kHz at the human eardrum (Whitehead *et al.*, 1997). The headphones incorporate foam covered eartips which insert comfortably into the auditory canal to block out external noise, and subjects used ER1-14A (large) or ER1-14B (small) eartips according to ear canal size.

### 2.5.4 Validation of VAS

Two tests were undertaken to ensure that a listener's DTF measurements were successful and that he or she constituted a valid subject for VAS experimentation. First, the listener was required to confirm subjectively that his or her individualised VAS stimuli were realistic and externalised. Second, the listener's ability to localise in VAS was compared to their ability to localise sounds in the free-field. Localisation performance was measured under the two conditions using the procedure described in section 2.4.1. A single free-field localisation test consisted of a series 150 ms broadband Gaussian white noise bursts (with 10 ms raised-cosine onset and offset ramps) presented randomly from 76 loudspeaker positions on the sphere. A single VAS test was identical except that stimuli were presented over earphones at corresponding 'virtual' positions. No feedback was provided, and each listener completed five repetitions of each kind of test.

Localisation performance summary statistics for each of the 10 subjects in free-field and VAS are listed in Table 2.1. The spherical correlation coefficient under each condition is shown, calculated on the basis of five repetitions at each of the 76 stimulus locations (i.e. 380 trials in total). It can be seen that all subjects localised with good accuracy in both free-field and VAS conditions, with SCCs ranging from

0.84 to 0.93. Also shown is the percentage of trials in which a cone of confusion error was made (see section 2.4.3). These errors occurred infrequently, with overall rates ranging from 0.3% to 4.7%.

**Table 2.1** Localisation performance summary statistics for each of the 10 subjects (S1-S10) in free-field and virtual auditory space (VAS). The spherical correlation coefficient (SCC) is calculated on the basis of five repetitions at each of the 76 stimulus locations (i.e. 380 trials in total). Also shown is the percentage of these trials in which a cone of confusion error was made. The calculation of these two statistics is described in section 2.4.3.

| | | S1 | S2 | S3 | S4 | S5 | S6 | S7 | S8 | S9 | S10 |
|---|---|---|---|---|---|---|---|---|---|---|---|
| SCC | free-field | 0.92 | 0.92 | 0.93 | 0.91 | 0.92 | 0.91 | 0.90 | 0.84 | 0.92 | 0.87 |
| | VAS | 0.85 | 0.92 | 0.90 | 0.91 | 0.90 | 0.88 | 0.88 | 0.85 | 0.87 | 0.86 |
| COC errors (%) | free-field | 1.8 | 1.3 | 1.1 | 0.3 | 1.8 | 2.1 | 3.7 | 3.4 | 1.3 | 3.5 |
| | VAS | 3.7 | 2.1 | 2.9 | 1.8 | 2.6 | 2.9 | 4.7 | 2.4 | 3.1 | 4.4 |

### 2.5.5 Spatial interpolation of DTFs

Importantly in the following experiments, it was often required that stimuli be presented at closely-spaced locations on the virtual sphere of space. However, in the DTF recording process only a discrete set of 393 locations was measured, distributed evenly on the sphere. Thus an interpolation procedure was applied to allow the description of the DTF at *any* location on the sphere based on the recorded DTF set. Interpolation was performed in the frequency domain using the method of spherical thin-plate splines (see Jin, 2001 for a mathematical description of this method). Comparisons of this approach to other approaches including linear interpolation and interpolation in the time domain have deemed this approach to be statistically and perceptually superior (Hartung *et al.*, 1999; Carlile and Leung, 2001).

In practice, the approach involved decomposing the magnitude components of the DTFs using principal components analysis (Carlile and Leung, 2001) to give rise to a series of principal components and weights. The weights were then interpolated using the spherical thin-plate spline, and DTFs were reconstructed from the principal components and interpolated weights. It has been confirmed in the laboratory that this procedure produces DTFs whose magnitude spectra are not statistically different from equivalent measured DTFs, and that virtual sound stimuli based on interpolated DTFs

are localised as well as those based on measured ones for the chosen DTF spacing (Carlile *et al.*, 2000).

# Chapter 3: Auditory two-point discrimination

## 3.1 Introduction

Despite an extensive body of research examining multi-source auditory environments, very few studies have attempted to measure systematically the spatial resolution of the auditory system for simultaneous stimuli. The following experiment was carried out to try to fill in this gap in the literature, and was based on the very simple two-point discrimination approach. This method has a long history in the visual and somatosensory systems, where it is of importance in both research and clinical domains.

Fine tactile tasks (such as reading braille) and the analysis of detailed visual images rely on accurate spatial discrimination in these modalities. Spatial discrimination is measured in touch by stimulating the skin gently with a pair of probes, and ascertaining the minimum separation at which the subject can feel that there are two stimuli. In vision, resolution thresholds are usually expressed as the smallest angular size at which a subject can discriminate the separation between critical elements of a stimulus pattern such as a pair of dots, a grating or a grid. In both of these situations, each stimulus is assumed to excite a restricted population of neurons at the various stages of the sensory pathway (receptors, relay nuclei, sensory cortex, etc.). This population displays a graded distribution of activity and it is the peak of this distribution that indicates the location of the stimulus on the body surface or in the visual field. For closely spaced stimuli presented concurrently, it is the interaction of these neural populations that determine the perceptual outcome. If the stimuli are so close that they excite almost the same set of neurons, then discrimination is impossible (Figure 3.1). However, it is proposed that lateral inhibition plays an important role in reducing the fusion of such excitatory regions, enabling remarkable spatial acuity in these systems (1 minute of arc in the fovea, 1-2 mm on the fingertips).

It was the aim of this experiment to use a similar approach to measure auditory spatial resolution for concurrent sources. However, it is important to recall that the

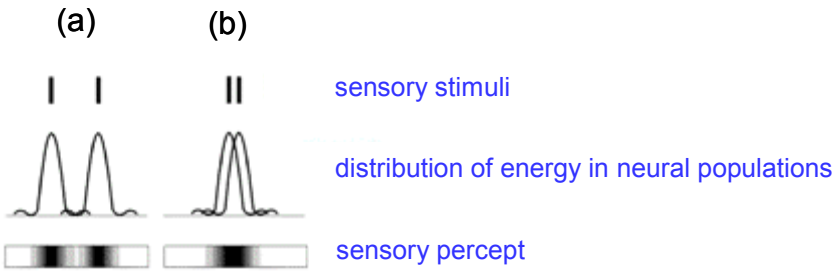



**Figure 3.1** Two-point discrimination theory. (a) The two vertical lines indicate the distance between two concurrent point stimuli, and the peaked functions represent the distribution of neuronal activity invoked by the two stimuli. The shaded bars at the bottom show that the two maxima give rise to two perceptual events. (b) When the stimuli are closer together (below the two-point discrimination threshold), the neuronal activity distributions overlap heavily. The shaded bar at the bottom indicates that the two maxima cannot be resolved and a single perceptual event occurs in this case.

mechanisms underlying spatial hearing differ in significant ways from those invoked in spatial vision or touch. As the auditory system does not possess a peripheral representation of space, and because it must compute location from a number of cues, we might predict the resolution of concurrent sources to be a more complex problem than in the other modalities. However if we assume that there are 'spatiotopic' maps at some level in the auditory system, there may be every reason to expect that the neural population effects described above are applicable.

## 3.2 Previous studies of auditory spatial resolution

Spatial resolution in audition has most commonly been measured using the minimum audible angle (MAA) and some discussion of this metric was given in a previous section (section 1.4.2). However this measure differs from traditional two-point discrimination measures, as it employs *sequential* rather than *simultaneous* pairs of stimuli. Only two studies were found that attempted to measure spatial resolution in a simultaneous sense.

Perrott (1984) defined the concurrent minimum angle (CMAA) as the threshold separation required to distinguish two concurrent sounds. He measured the CMAA for sources distributed in a horizontal plane and reported it to be larger than the angle required to determine the direction of displacement of two sequential sounds (MAA). Perrott presented pairs of tones of different frequency, and asked subjects to judge the relative location of the pair by indicating whether the higher tone was to the

left or right of the lower tone. Using a criterion of 75% correct, he measured CMAAs and noted a significant effect of azimuth, with CMAAs of 4 - 10° at the front increasing to 30 - 45° at a lateral displacement of 67°. Divenyi and Oliver (1989) extended this work to examine more complex sounds including pairs of amplitude-modulated tones, pairs of frequency-modulated tones, and pairs of sinusoidal carriers (differing in frequency) frequency-modulated by a common broadband noise source. Stimuli were presented in non-individualised virtual auditory space and simulated locations were on the frontal horizontal plane. These authors found a similar threshold increase for stimuli located at 80° as compared to 0° azimuth.

In these studies, the concurrent stimuli differed in their frequency content and so binaural effects could not be isolated from pitch effects. In fact Perrott (1984) reported that CMAAs increased if the frequency difference between the two sources was reduced, indicating that the two parameters are confounded. Divenyi and Oliver (1989) also noted that, for their more broadband stimuli, spectral overlap was detrimental to resolution. In the present experiment, this sort of confound was avoided by presenting stimuli with the same source content. It was considered important that the task remain a purely spatial one in order to faithfully measure two-point discrimination.

Another limitation of the studies cited above, and indeed most studies of multiple source perception, is that they have only examined separation in the horizontal plane. In the present study, one objective was to examine sound sources coming from many directions around the listener, displaced both horizontally and vertically. Importantly, broadband noise stimuli were chosen because they are well localised in all dimensions (Carlile *et al.*, 1997) as opposed to narrowband stimuli which are poorly localised in elevation and in the front-back dimension (Butler, 1986; Middlebrooks, 1992). Effectively, in previous experiments, the bandwidth required for accurate localisation had been sacrificed for the sake of gaining an adequate frequency separation of two sources.

## 3.3 Approach

In the present study the competing sounds were broadband noises, and subjects were asked to simply indicate whether they perceived sound arriving from a single location

or from two distinct locations (this is called 'separation' throughout the thesis, and has also been called 'detection of spatial separateness', Noble *et al.*, 1997). However, our 'simple' task is quite difficult for two reasons. Firstly, the two sources presented are identical in their long term spectrum and temporal characteristics and so cannot be segregated on the basis of their content or identity. Secondly, as these sounds are broadband, they both produce global activation along the basilar membrane and thus a complete sharing of receptors must take place. Separation cannot commence at the periphery as is the case with differing tones, which are processed in distinct frequency channels, but must rely on more central processing of spatial cues. It is not entirely clear what acoustical cues (for localisation and/or separation) are available in the sum of two such stimuli.

Experiment 1 aimed to examine the ability of listeners to resolve the simultaneous pair and compared this ability in different spatial regions around the listener. By choosing specific spatial configurations it was possible to gain some insight into the contribution of different localisation cues to performance. The results of Experiment 1 prompted a series of further experiments that examined the potential role of ITD in some detail. Recall that the ITD can be conveyed in both high and low frequency channels as well as in the onset and offset of a transient sound (section 1.3.1). In the following experiments, separation of concurrent noise sources was examined upon removal of high-frequency content (Experiment 2a), low-frequency content (Experiment 2b), and onset/offset ITDs (Experiment 2c). In Experiments 3a and 3b the effect of removing these cues in combination was examined.

## 3.4 Experimental methods

### 3.4.1 Subjects and task

Four subjects (S1-S4) participated in the experiments. All subjects completed each of the experiments except for Experiment 2a which was completed only by S1. Experiments were carried out in the soundproof room and stimuli were presented in virtual auditory space (Chapter 2). Subjects S2-S4 completed 34 listening tests in total (approximately 8.5 hours listening time) and subject S1 completed 44 listening tests (approximately 11 hours listening time). Generally subjects completed 2-3 tests in a

single session with short breaks in between, and the frequency of sessions ranged from 2-8 per week depending on the subject.

The task took the form of a single-interval forced choice procedure (see section 3.4.4 for a short discussion of this approach). Subjects were presented with a stimulus and were asked to indicate, by pressing one of two buttons, whether they perceived sound to be coming from one or two source locations. The two source locations were termed the 'reference' and the 'test' locations as described in the next section. Subjects were familiarised with the stimuli and task before the commencement of testing (see section 3.4.4).

### 3.4.2 Stimulus configurations

The general technique for generating virtual stimuli was as described previously (section 2.5), but it was extended to simulate simultaneous sound sources in different locations. This involved simply generating the two stimuli independently and adding the two resultant signals for each ear linearly. For 'zero separation' stimuli, the same procedure was followed but the two DTF pairs used were identical. All stimuli were 150 ms in duration and after summing produced a sensation level of approximately 50 dB. Details of the stimulus manipulations used in the different experiments are given in the individual sections.

Stimulus locations are described in this chapter using azimuth and elevation with respect to a single-pole co-ordinate system (section 2.2). For Experiments 1, 2a, 2b, and 2c testing took place at five reference locations on the 0° elevation horizontal plane: 0°, 22.5°, 45°, 67.5° and 90° azimuth (Figure 3.2a). In each trial, one stimulus in the concurrent pair was presented from a reference location and the other from a test location displaced in either azimuth (horizontal separation) or elevation (vertical separation) or from the same location (Figure 3.2b). Note that binaural cues change with horizontal separation (maximum rate-of-change at the front, minimum rate-of-change at the side) and also with vertical separation using the single-pole co-ordinate system (maximum rate-of-change at the side and *no* change at the front). For each reference location, 15 test locations were chosen on the basis of preliminary testing. For vertical separation, the testing range was the same for each of the five locations, and spanned the entire available range of elevations: from -45° to 90° (directly

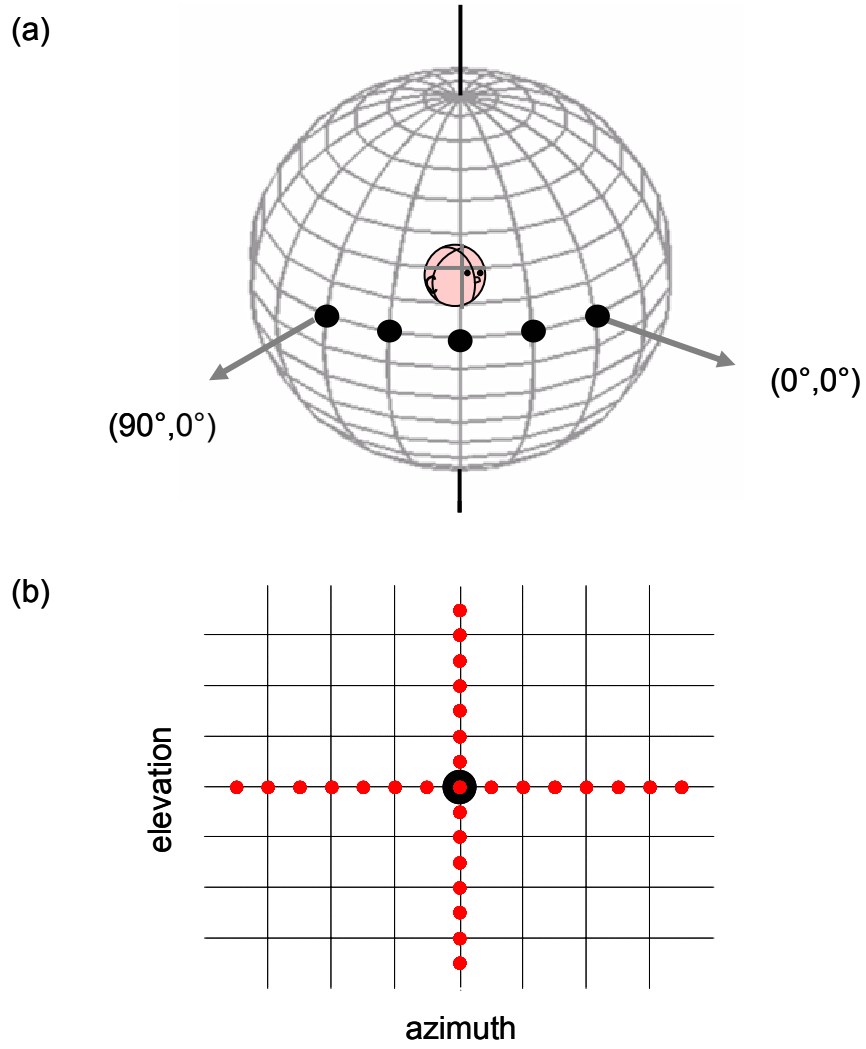



**Figure 3.2** Stimulus configurations, described using the azimuth/elevation co-ordinate system. (a) Five reference locations were employed in the right frontal quadrant on the 0° elevation plane: azimuths 0°, 22.5°, 45°, 67.5° and 90° (shown as black dots). (b) To create stimulus location pairs, each of these reference locations was paired with 15 test locations distributed in azimuth and elevation (red dots, see section 3.4.2 for details).

overhead). Ranges for horizontal separation varied with location as required to cover a suitable range: ± 21° for 0° azimuth; ± 32° for 22.5° azimuth; ±42° for 45° azimuth; ±53° for 67.5° azimuth; ±63° for 90° azimuth. All reference location/separation combinations were presented 10 times each in a random order, giving 1500 trials in an experimental block (5 reference locations, 2 directions, 15 test separations, 10 repetitions). These were broken down into 10 tests of 150 trials.

For Experiments 3a and 3b testing only took place at the most frontal reference location (0° azimuth). The same test separations were used for this location as in the other experiments (vertical: 15 values between –45° and 90° elevation; horizontal: 15 values between –21° and 21° azimuth). These shorter experiments each comprised 300 trials (two tests of 150 trials).

### 3.4.3 Data analysis

Results were analysed on a subject-by-subject basis for each block of experiments in turn. For each reference location, responses to the different stimulus configurations were sorted and psychophysical curves for horizontal and vertical separation were plotted. These illustrate, for each test separation, the percentage of times (out of 10 repetitions) that the subject responded that he/she perceived two sources.

### 3.4.4 A note on response criteria, training and controls

In establishing the experimental protocol for examining the spatial perception of concurrent stimuli, the widely-used two-interval discrimination task was piloted; it required subjects to choose which interval contained two sources and which contained only one. However, this approach was quickly rejected because it was clear that comparisons could be made on the basis of a number of available cues (such as timbral changes), and not necessarily the ones of interest in this study (spatial cues). The single-interval paradigm was thus adopted in order to best measure the effects of spatial separation on the perception of concurrent sources. When presented with a particular stimulus, subjects were required to respond as to whether they perceived one or two source locations. Although this approach did not *ensure* that responses were related to a clear percept of two sources, it did encourage listeners to use cues that were (subjectively) spatial in nature. Note that while this task does not represent a true discrimination task, the term ‘two-point discrimination’ is maintained during the thesis to reflect the relationship of this task to the classic visual task of the same name.

One difficulty associated with the single-interval subjective task is that response biases (such as pressing one button more often than the other) cannot be effectively removed as they can in a discrimination task. Only if this bias is consistent

within a subject can comparisons across conditions be made with confidence. Another potential problem with the one-interval task is its subjectivity: the experimenter must rely on subjects to adopt a criterion for responding and to maintain this criterion throughout testing. These potential difficulties were dealt with in two ways. Firstly, subjects underwent a small amount of training to stabilise their performance before commencement of data collection. Each subject was run through two tests of the format described in section 3.4.1 allowing him/her to become familiar with the stimulus and task and to establish a comfortable criterion. Secondly, throughout an experimental block, a specific repeated control was included in each test to ensure that a particular subject's bias and criteria did not fluctuate. This control consisted of a small set of 30 trials, which were interleaved with the 150 experimental trials. These were five repeated trials at six horizontal separation values from the 0° reference location set (separations of 0, 3, 6, 9, 12, 15°). Responses to these trials provided a measure of response bias and criterion for each test. At the end of an experimental block, the control sets were examined to confirm that they were stable across testing. If a set deviated greatly from the rest, as determined by the experimenters on the basis of informal monitoring, the test from which it came was repeated. Only two tests had to be repeated on this basis (in one subject).

Despite the good within-subject consistency, the fact that response bias and criteria varied across individuals meant that it was difficult to pool responses across the population. Thus for this study the emphasis is not on the absolute value of responses, but rather the pattern of performance of each subject across the various stimuli and test locations.

## 3.5 Experiment 1: Broadband stimuli

### 3.5.1 Stimuli

Experiment 1 was conducted to examine the ability of subjects to separate concurrent broadband sounds. Random noises containing frequencies from 300 Hz - 16 kHz were used (as for localisation testing, see section 2.4.1). For a given stimulus, two independent noises were filtered with appropriate DTFs and ramped by applying a

raised cosine to the first and last 10 ms. The two binaural signals were then added to create concurrent stimuli as depicted in Figure 3.3a.

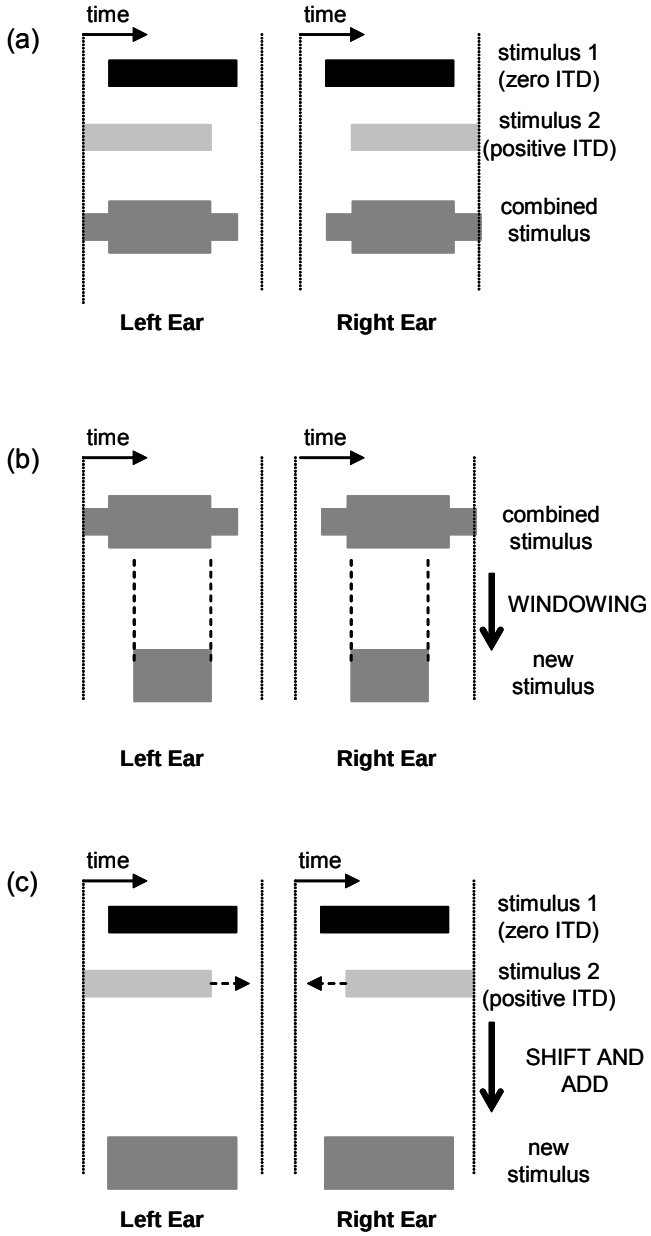



**Figure 3.3** Temporal features of stimulus manipulations used in Experiment 1, 2c, and 3b. The examples given are for a reference stimulus located frontally and a test stimulus at a single lateral location. Left and right ear stimuli are shown in left and right columns. (a) Experiment 1: To create a concurrent pair, the reference stimulus (zero ITD – black bars) was added to a laterally displaced sound (positive ITD - light gray bars) to create a new stimulus. The ITD difference created two onsets and offsets in each ear, as well as a phase difference in the ongoing portion of the signals. (b) Experiment 2c: The combined stimulus just described was windowed identically in each ear to remove the onset and offset cues. The ongoing phase differences remained in the signal. (c) Experiment 3b: To remove all ITD differences between the two stimuli, the laterally displaced stimulus (light grey bars) was shifted in time in each ear before adding to align it to the zero ITD stimulus. This removed both onset/offset and ongoing ITD differences, but individual spectral characteristics were preserved.

### 3.5.2 Results

Psychophysical curves for the first experiment are shown in Figure 3.4. Data from individual subjects (S1 to S4) are shown in parts (a) to (d) and mean data are shown in part (e). The left hand panels show data for concurrent sources that were separated in azimuth (horizontally) about the reference location and the five subplots in a panel show data for each of the five reference locations. The abscissa shows the separation of the test stimulus with respect to the reference, where negative values indicate leftward separation and positive values indicate rightward separation. Plotted on the ordinate is the percentage of trials (out of 10) where the subject reported that he/she perceived two sources in the presentation.

A feature that is evident for all subjects (and the mean data) is that the psychophysical curves for horizontal separation broaden as the reference position is displaced more laterally, indicating a decrease in spatial sensitivity with increasing laterality of the sources (left panels, Figure 3.4). For the most frontal reference location (0º, 0º), 100% response rate was reached with much smaller separation angles than for the most lateral location (90º, 0º). All subjects showed a similar pattern of results in this experiment, although there were marked differences in overall response levels. For example, S4 had a strong tendency to respond to perceiving both sources, as indicated by the relatively narrow troughs (Figure 3.4d), whilst S1 tended to give more conservative responses (Figure 3.4a). This individuality is examined in section 3.8.1. A feature of the data for all subjects is an asymmetry in the psychophysical curves for 67.5º reference location with a drop in response rate as test stimuli were moved towards the back (positive separation). An explanation for this is given in the discussion of this experiment.

The right hand panels of Figure 3.4 show data plotted in a similar way, but for sources separated vertically about the reference location. Again each of the five reference locations are plotted in separate subplots, and the axes are as described above, except that the separation was in elevation (vertical), with negative values indicating downward separation and positive values indicating upward separation. In these data, the opposite trend can be seen, in that the curves appear to get slightly sharper at the more lateral locations. By far the most striking feature of these data is the frontal midline location. Subjects S1 and S3 show a completely flat psychometric

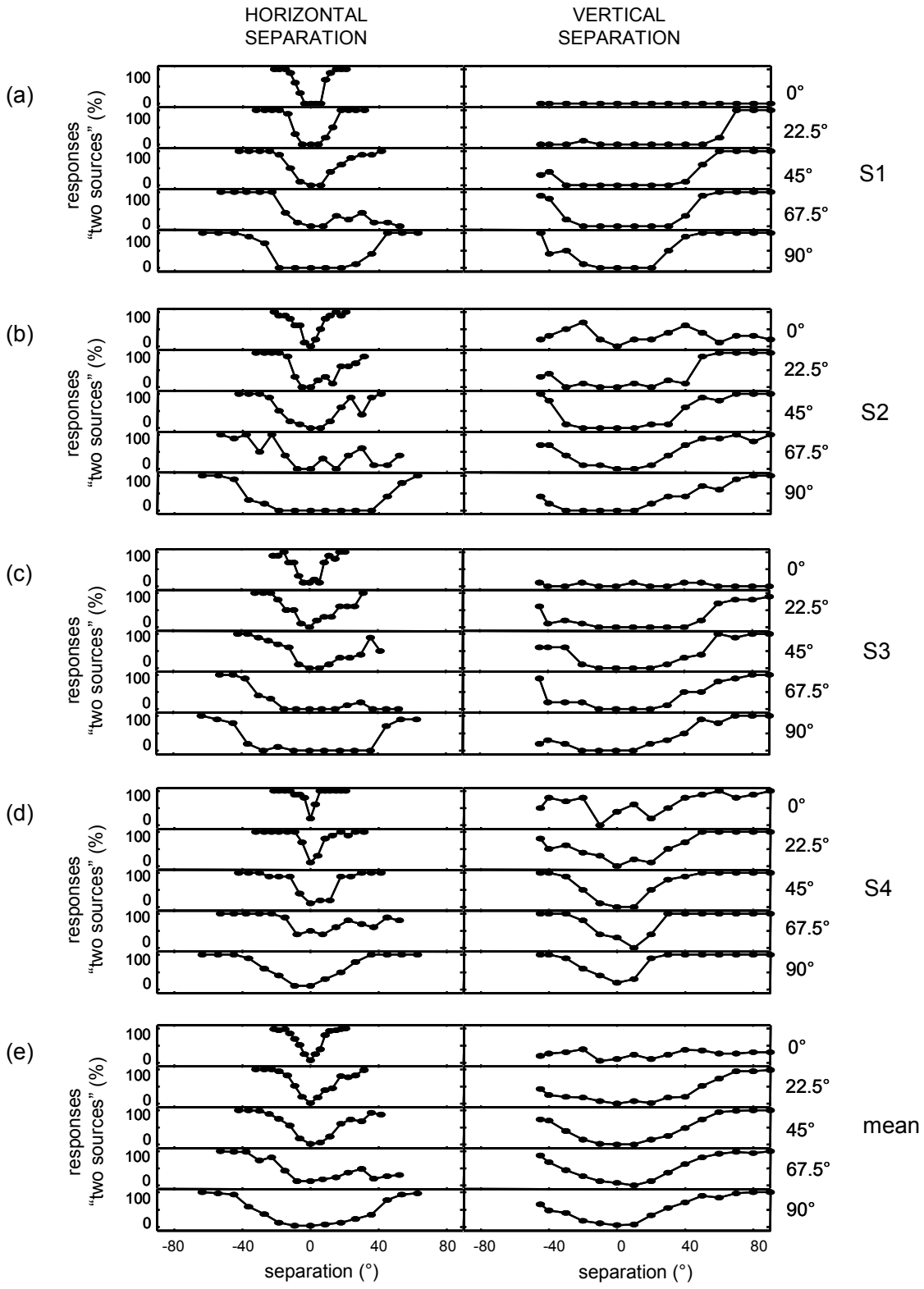



**Figure 3.4** Psychophysical curves for Experiment 1. (a) to (d) show data for subjects S1, S2, S3, and S4 respectively and (e) is mean data. The left hand column shows horizontal separation data and the right hand column shows vertical separation data. The five subplots in each panel show data for each of the five reference locations (top to bottom: 0°, 22.5°, 45°, 67.5° and 90° azimuth). For the given test separations (abscissa), the curves show the percentage of trials in which the subject perceived two sources.

function indicating that they never (S1) or rarely (S3) perceived both sources in the pair. S2 gave more positive responses to the stimulus pairs overall but the curve is still relatively flat, indicating that responses were not related to separation. Responses from S4 were inconsistent, but he seemed to be able to resolve the sources at large positive separation values.

### 3.5.3 Discussion

**Binaural trends in the data**

The trends seen in the data as the spatial location of testing was varied point strongly to a role for binaural cues in this particular task. The data obtained for horizontal separation on the 0° elevation plane are certainly consistent with this idea. As a result of the position of the two ears on this plane, binaural cues change approximately as a sine function of azimuth, with the rate-of-change being maximal at 0° azimuth and decreasing towards 90° azimuth. As a result, performance in spatial discrimination tasks (Mills, 1972) and absolute localisation (Carlile *et al.*, 1997) become poorer in the horizontal dimension as azimuth increases. In addition, it follows that if separation is dependent on differences in the binaural cues between a pair of stimuli, then the *angle* required to achieve adequate binaural separation would increase with laterality. This is seen in the present data, where the angle of separation required to reach a given response rate increased at the more lateral positions. A special case can be seen at the 67.5° reference position when the test stimuli were separated towards the back. The value of the ITD and ILD for the test stimulus increases as the test location is moved towards 90° azimuth (separations of 9° and 18°) but then *decreases* back towards the reference value when the test location passes 90° (separations of 27° and 36°). When the test location is 112.5° azimuth (separation of 45°), this is the reflection of 67.5° azimuth about the interaural axis and so the binaural cues for the test and reference location are near to identical. Thus these data support the notion that the task is dependent on binaural differences between the two sources.

The pattern of responses seen with vertical separation also suggests that the rate-of-change of the binaural cues is, at least in part, responsible for performance. At the most lateral reference location (90° azimuth), vertical separation of a concurrent stimulus causes a greater change in binaural cues than at more frontal locations, as the

displacement is towards the far ear and away from the near ear. Furthermore, at this location, a given separation in the vertical direction changes the binaural cues by approximately the same amount as the same separation in the horizontal direction. Consistent with this, the psychophysical curves at 90° azimuth are similar for horizontal and vertical separation. Perhaps the clearest demonstration that binaural cues are important for this task is the fact that when stimuli were separated along the vertical midline, subjects generally perceived only one source (or were confused). In other words, given only one binaural value, the auditory system assumes one source.

**Estimates of angular threshold for separation**

In order to summarise the data, perceptual thresholds were estimated from the psychophysical curves for all subjects in Experiment 1. To facilitate this estimation, standard psychometric curves were fit to the raw data using Probit analysis (Finney, 1971). Perceptual threshold was defined as the separation value where the by two source locations were reported to be perceived in 75% of cases. As 'positive' and 'negative' separation values were tested (corresponding to left and right, or up and down) an upper and lower threshold was obtained separately from the two halves of each curve and threshold was taken as the mean of these. It was not possible to obtain a value in some cases, where performance did not reach the 75% level within the range of testing. Figure 3.5a illustrates these angular thresholds (blue bars). The five panels represent the individual subjects and the mean across subjects, and the left and right halves of each panel show the horizontal and vertical thresholds respectively.

These summary plots demonstrate for horizontal separation (Figure 3.5a, left panels) the trend already discussed: thresholds increase with increasing lateral position. Consistent with this, Perrott (1984) reported that the concurrent minimum audible angle of tones increased at lateral positions, and Divenyi and Oliver (1989) found a similar effect for amplitude- and frequency-modulated tones. The mean thresholds for horizontal separation (Figure 3.5a, lower left panel) are in reasonable agreement with the CMAAs reported by Perrott (1984). The subjects in the present experiment required 9.2° on average to resolve a concurrent pair at a reference azimuth of 0° (Perrott's subjects: 4-10°) and 26.2° at a reference azimuth of 67.5° (Perrott's subjects: 30 - 45° at a reference azimuth of 67°).

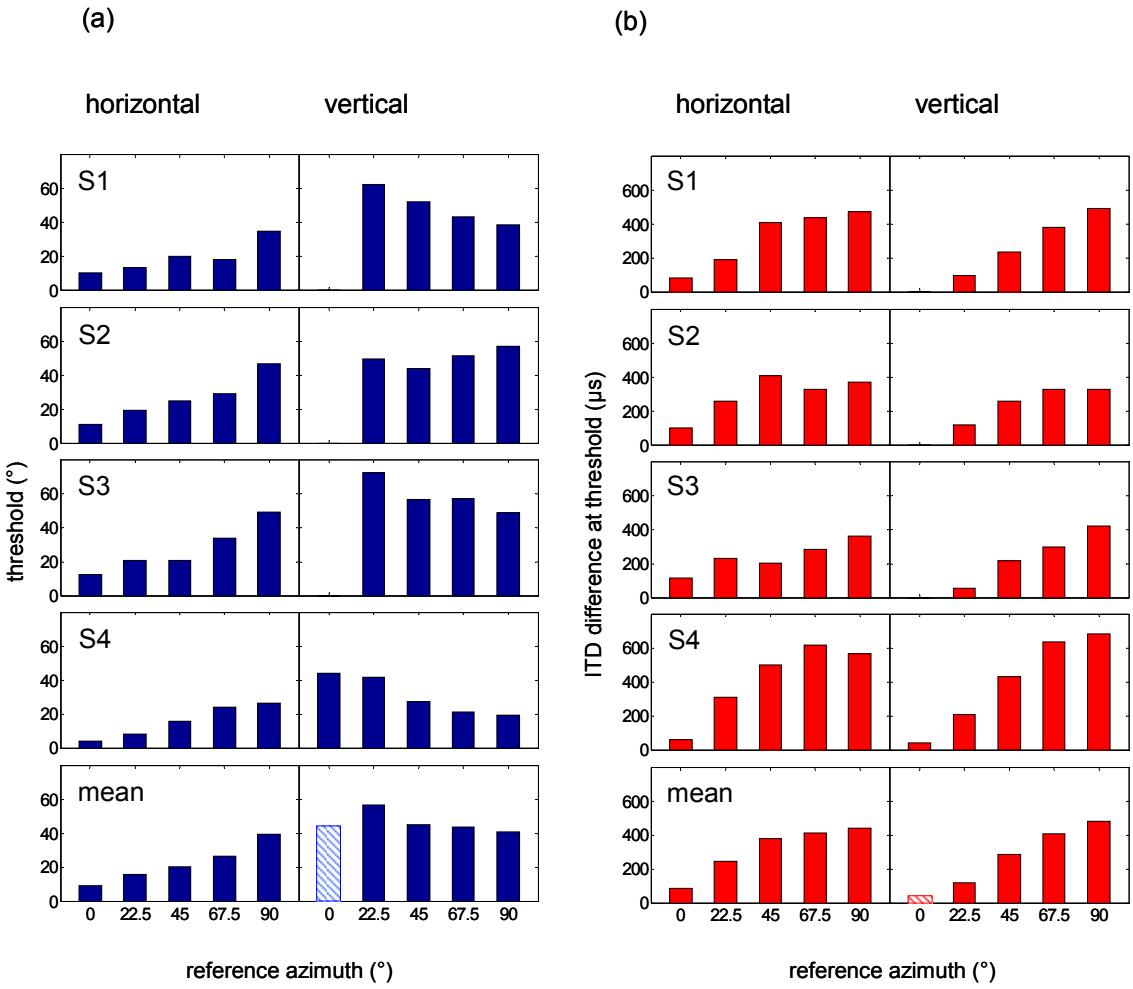



**Figure 3.5** (a) Perceptual thresholds for separation (calculated from the psychophysical curves in Figure 3.4). The five rows show data for S1, S2, S3, S4 and the mean data as labelled. The left and right halves of each panel show data for horizontal and vertical separation respectively. The five blue bars represent perceptual thresholds at each of the five reference locations. (b) ITD differences at perceptual threshold. The five rows show data for S1, S2, S3, S4 and the mean data as labelled. The left and right halves of each panel show data for horizontal and vertical separation respectively. The five red bars represent ITD differences at each of the five reference locations, and are calculated from the equivalent blue bars in Figure 3.5a. Hashed bars indicate means calculated from only one subject (S4).

For vertical separation (Figure 3.5a, right panels) there are more individual differences in thresholds, and overall the thresholds are large (40.9º at best, at 90º azimuth). For 0° reference azimuth, a threshold could only be obtained for S4 (43.85º). For the other subjects threshold was not reached within the testing range (the largest separation tested was 90º). The implication here is that spectral cues are not sufficient under these circumstances to indicate the presence of the distinct sources. This may, at first, seem surprising because spectral cues are responsible for an accurate ability to localise single sources on the vertical midline: average errors are approximately 5º at

(0º, 0º) for broadband noise (Carlile *et al.*, 1997; Butler *et al.*, 1990). In addition, Perrott and Saberi (1990) reported that subjects could detect a change in location between sequentially presented click trains when they were separated by 3.5º along the vertical midline. Clearly, however, when a pair of broadband sounds is presented concurrently, as in the current experiment, the individual spectral cues become less useful as they sum at each ear.

For each case where a threshold value could be obtained, it was possible to express the angle in terms of the *ITD difference* between the two sources. This enables a preliminary examination of the idea that binaural cues are indeed driving the response patterns. ITD differences for each individual were calculated by subtracting the ITDs of the reference and test positions at threshold. These ITDs were extracted from the impulse responses of each subject (measured for the creation of virtual auditory space see section 2.5), by cross-correlating the left and right ear signals and determining the interaural delay giving maximum correlation. Figure 3.5b (red bars) shows the average ITD difference calculated from each of the perceptual thresholds plotted in Figure 3.5a. Here it can be seen that, when expressed as a binaural difference, the thresholds for horizontal and vertical separation at a particular reference location are similar. This is good evidence that response patterns are related to binaural thresholds and that these are unique to (but vary across) spatial locations. The progressive increase in ITD thresholds as azimuth increases is discussed further in section 3.8.2.

**A unique subject**

The idea that binaural differences are required for resolution of concurrent pairs explains well the general finding that performance was poor on the median plane (where binaural cues are relatively constant as a result of the symmetrical positioning of the two ears). Recall however that subject S4 did seem to have some ability to resolve the concurrent pair when they were separated along the vertical midline (Figure 3.4d, right hand upper panel). An examination of this subject's head-related transfer functions revealed a marked asymmetry (due to an asymmetry in the position of the ears on the head) resulting in elevation-dependent binaural cues along the 0° azimuth vertical dimension. As a measure of binaural changes, the ITD was calculated from this subject's measured impulse responses at 10° steps between -40° and 90° elevation. Figure 3.6 shows the ITD function of this subject (red line) as compared to

the other three subjects. S4's asymmetry is apparent in that the ITD increases as a function of elevation. Specifically, the difference in ITD between elevations of 0° and 90° was measured to be 75 μs (compared to the mean value of 12.5 μs for the other three subjects) and this difference is indeed large enough to be a useful cue for separation at the frontal location (see Figure 3.5b and section 3.8.2).

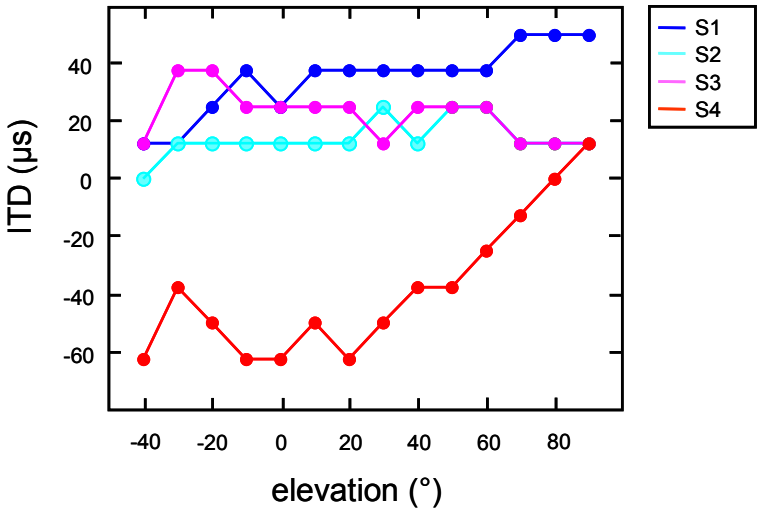



**Figure 3.6** ITD functions along the median vertical plane for the four subjects. For elevations of –40° to 90° (abscissa), ITD values estimated from impulse responses are plotted (ordinate). It can be seen that S4 (red line) shows a marked asymmetry; the ITD increases as a function of elevation.

## 3.6 Experiment 2: Effect of removing single components of the ITD

### 3.6.1 Stimuli

Experiment 2 was conducted in the same way as Experiment 1 but with stimulus manipulations to remove components of the ITD cue. It was predicted that the results of Experiment 1 could be explained by dominant low-frequency ITD cues, and Experiment 2a was aimed at confirming this prediction. This experiment was completed by S1 only. Noise stimuli were low-pass filtered at 1.2 kHz before convolving with DTFs. Filtering was performed in the Fourier domain by zeroing the

magnitude of all components above 1.2 kHz and taking the inverse transform to return to a time domain waveform. In Experiment 2b, which was completed by all subjects, stimuli were high-pass filtered at 2 kHz in order to remove this dominant low-frequency ongoing timing information. Filtering was again performed in the Fourier domain by zeroing the magnitude of all components below 2 kHz. In Experiment 2c, again completed by all subjects, the aim was to examine the influence of onset and offset ITD information. This was achieved by adding the two DTF-filtered noises together and then windowing the resultant left and right ear stimuli such that they ramped on and off at exactly the same points in time. Specifically, 2 ms was trimmed from each end of the combined stimuli, which was enough to ensure the removal of the two onset and offset cues that were previously available (Figure 3.3b). Because 4 ms of signal was lost in the windowing process, the original sounds were generated with duration 154 ms to give rise to a final 150 ms signal. Technically, this manipulation sets the onset and offset ITDs to zero, as opposed to removing them. Nonetheless, it provides the opportunity to observe the effects of removing the location-dependent variation in this cue.

### 3.6.2 Results

S1's results from Experiment 2a, where stimuli were low-pass filtered, are shown in Figure 3.7. Data are shown for low-passed stimuli (solid lines) as compared to the broadband data described in Experiment 1 (dotted lines). The same set of separations was tested in this experiment (and those following) as in Experiment 1, however, the symbols have been omitted from the figures for clarity. The curves represent the percentage of responses in which the subject responded that both sources in the concurrent pair were perceived. It can be seen that this subject performed similarly using low-passed stimuli as when using broadband stimuli. There are, however, deviations in the curves for some configurations, and in general these indicate a weaker tendency to respond to hearing both sources in the low-pass condition.

Figure 3.8 shows data for all subjects in Experiments 2b and 2c. In each panel, results for high-passed stimuli (solid lines) and stimuli with onsets and offsets removed (dashed lines) are plotted along with the broadband data described in Experiment 1 (dotted lines). The curves represent the percentage of responses in

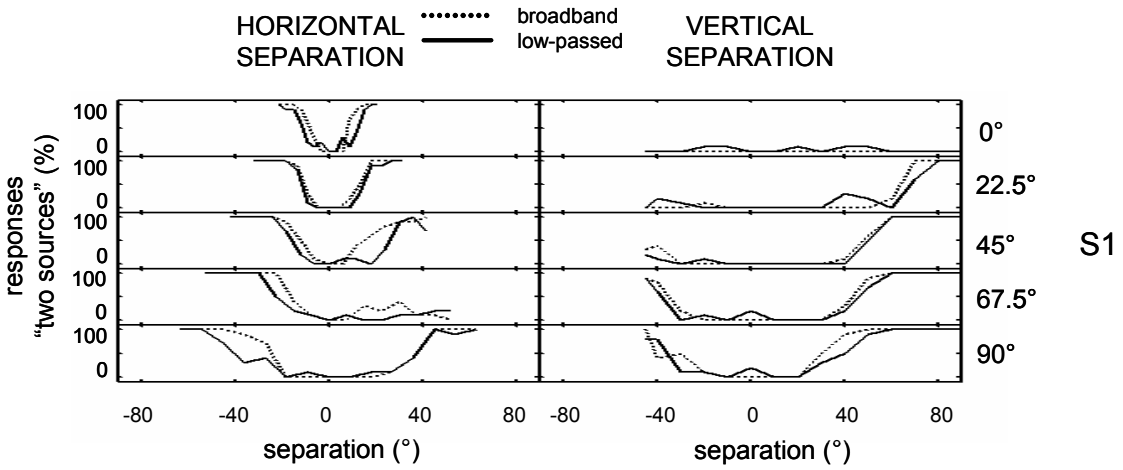



**Figure 3.7** Psychophysical curves for S1 in Experiment 2a. Results are shown for broadband stimuli (dotted lines; same as Figure 3.4a) and stimuli low-pass filtered at 1.2 kHz (solid lines). All other details as for Figure 3.4.

which the subjects responded that they heard both sources in the concurrent pair. Results for subjects 1 to 4 are shown in parts (a) to (d) and the mean in part (e).

It can be seen that for S4, behaviour did not change substantially as a result of the high-pass filtering. The curves (solid lines) for this subject generally agree with the broadband condition (dotted lines), with minor deviations. For S2 this is generally true also, although some larger deviations can be seen particularly for horizontal separation at the more lateral locations. For S1 and S3, the high-pass filtering had a more obvious impact on performance, but again only at the more lateral reference locations. Here the psychophysical curves became flatter, and 100% separation rate was reached far less often than in the broadband condition. In an examination of the mean data (Figure 3.8e) it appears that the front most reference location (0° azimuth) was not affected by the high-pass filtering. However, some flattening of the curves is apparent at more lateral locations (especially 67.5 and 90° azimuth) representing a decrease in sensitivity to the task.

It can also be seen for each individual that there was no substantial effect of removing the onset and offset ITD cues on performance; the curves (dashed lines) are generally in agreement with the broadband condition (dotted lines). This is true for both horizontal (left panel) and vertical (right panel) separation and is confirmed in the mean data.

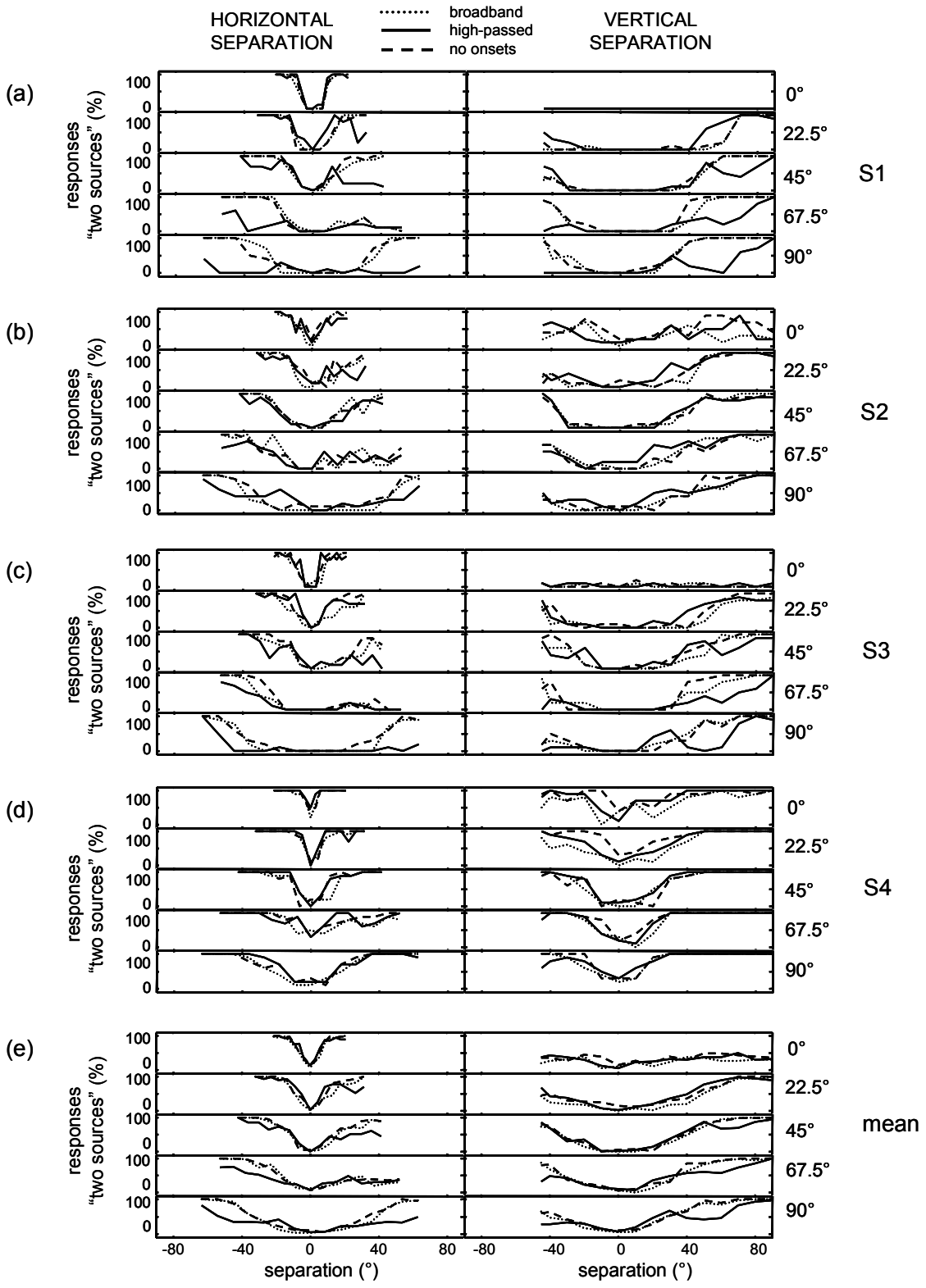



**Figure 3.8** Psychophysical curves for Experiments 2b and 2c. (a) to (d) show data for subjects S1, S2, S3, and S4 respectively and (e) is mean data. Results are shown for broadband stimuli (dotted lines; same as Figure 3.4a), stimuli high-pass filtered at 2 kHz (solid lines), and stimuli with onset/offset cues removed (dashed lines). All other details as for Figure 3.4.

### 3.6.3 Discussion

It was interesting to find that the separation of concurrent sounds at the frontal location was not impaired by high-pass filtering the signals. It seems that in this situation, the ongoing low-frequency ITDs are redundant, and the perception is maintained by high-frequency acoustical cues in this case. It has been shown previously that high-pass filtering does not significantly affect the accuracy of localisation of *single-source* stimuli (Butler and Humanski, 1992; Carlile and Delaney, 1999), and the present findings suggest that high-frequency cues can also be sufficient for identifying that more than one location is present in a *multiple-source* stimulus.

In some subjects, high-pass filtering was seen to disrupt performance at the more lateral locations. Indeed the task was performed more poorly at these locations in the broadband condition, but with high-passed stimuli performance in some subjects dropped severely. In fact for one subject the response rate dropped to around zero for *any* stimulus pair presented at the most lateral location (S1, Figure 3.8a). It is possible that this is related to ITD discrimination thresholds in this region, and this is discussed further in section 3.8.2.

Another finding from Experiment 2 was that the presence of ITDs at the onset and offset of the stimulus was not necessary to match the performance of subjects seen in Experiment 1. This is somewhat surprising considering that many neurons in the auditory pathway respond preferentially to the onset of a sound stimulus (Pickles, 1988). It is also curious because the onset cue is available and consistent across *all* frequency channels. Indeed for very short impulsive sounds such as clicks, it is the only cue available, and is crucial for lateralisation (Tobias and Schubert, 1959). For longer duration stimuli under anechoic conditions, the ongoing phase cue is likely to be a more reliable cue because it can be integrated over time to improve the estimate. It seems that the transient offset and onset disparities are redundant in the presence of this very robust cue. A similar redundancy of onset ITD information has been shown for the lateralisation of ongoing noise bursts presented over headphones (Tobias and Schubert, 1959).

# 3.7 Experiment 3: Effect of removing ITD components in combination

## 3.7.1 Stimuli

Experiment 3 combined some of the stimulus manipulations of Experiment 2 in order to assess potential redundancies in the ITD cue set. These conditions were tested only at the frontal location (0°, 0°), as results from the previous conditions had indicated that this location was the only one where no degradation was seen using the ITD manipulations individually.

Experiment 3a was a combination of Experiments 2b and 2c, where stimuli were high-pass filtered at 2 kHz *and* left and right ear signals were each windowed as just described. In this way, onset and offset ITDs as well as ongoing low-frequency ITD information was removed. This allowed an examination of the usefulness of *high-frequency* information for this separation task, preserving only high-frequency ongoing ITD as well as spectral information.

In the final experiment (Experiment 3b) the aim was to remove *all* ITD information. To achieve this, the paired spatial stimuli were generated as for Experiment 1. Following this, the left and right ear signals representing the test stimulus were shifted in time (if required) such that their onsets matched those of the left and right ear signals of a stimulus corresponding to location (0°, 0°). The signals were then ramped and added (Figure 3.3c). The result of this manipulation was that the onsets and offsets of the stimulus pair coincided, and the ongoing time differences too were matched. Thus the timing information in isolation corresponded to only one location in space, but the spectral and interaural level information from each individual source was still contained in the composite signal.

## 3.7.2 Results

Each panel in Figure 3.9 contains data from the previous experiments, as well as data from Experiment 3, for horizontal separation at the frontal location only. Individual data for subjects 1 to 4 are shown (Figure 3.9a, b, c, d) as well as the mean data

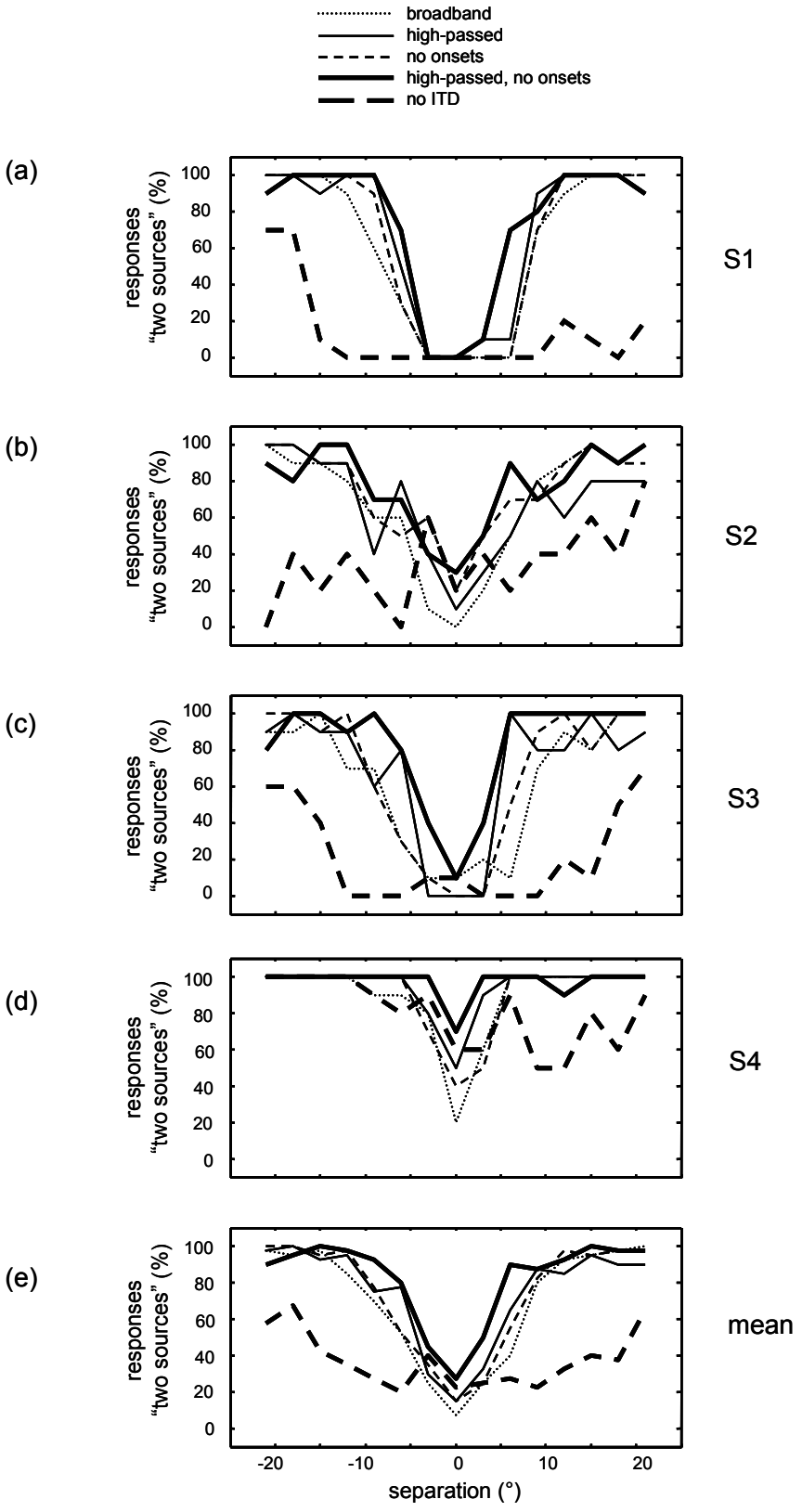



**Figure 3.9** Psychophysical curves for Experiments 3a and 3b (horizontal separation at the frontal location only). (a) to (d) show data for subjects 1 to 4 respectively and (e) shows mean data. Data shown again for Experiments 1, 2b and 2c (thin dotted lines, thin solid lines and thin dashed lines) as well as Experiment 3a (high-pass filtered and onset/offset cues removed; thick solid lines) and Experiment 3b (all ITD information removed; thick dashed lines). For the given test separations (abscissa), the curves show the percentage of trials in which the subject perceived two sources (ordinate).

(Figure 3.9e). Experiment 1 (dotted lines), Experiment 2a (thin solid lines), and Experiment 2b (thin dashed lines) have already been discussed, and again the relatively narrow troughs can be seen, indicating a good level of performance in this testing range. The thicker lines show data from Experiment 3: high-pass filtered stimuli with onset/offset cues removed (thick solid lines) and stimuli with all ITD cues removed (thick dashed lines).

It can be seen that the curves from Experiment 3a (thick solid lines) are very similar to the previous data, indicating that the combination of stimulus manipulations from Experiment 2 did not degrade performance at this location. In fact, it seems that in this condition subjects tended to report *more often* that they perceived two sources.

Psychophysical curves for Experiment 3b (thick dashed lines) show large individual differences, but it can clearly be seen that the consistent performance obtained in all other conditions was disrupted in this condition. Across all spatial configurations, the number of trials where both stimuli in a concurrent pair were reported was lower than in the other conditions. This is true for the individual subjects (Figure 3.9a, b, c, d) and for the mean (Figure 3.9e). Even for the largest separations, response rate never reached 100% for S1, S2 and S3. S4 showed a marked performance asymmetry, displaying a reasonable ability to distinguish the two sources with leftward separation, but a disrupted performance with rightward separation. This is in line with the asymmetry of this subject's auditory periphery, discussed previously in section 3.5.3.

### 3.7.3 Discussion

Results from Experiment 1 suggested that binaural differences drove the separation of concurrent broadband noise sources in the 4 subjects examined. Experiment 2a confirmed (in one subject) that low-frequency information was sufficient to maintain performance observed in the broadband condition. However in Experiment 2b it was found that low-frequency phase information was not crucial to perform this task, although it aided performance at the more lateral locations. Experiment 2c showed that onset and offset ITD information was not critical and did not even appear to contribute to the performance seen in Experiment 1. As there was a possibility of redundancy between these latter two sets of cues, they were both eliminated in

Experiment 3a to assess their combined importance. At the frontal location, performance was not disrupted under these conditions. Remarkably, subjects could determine the number of sources in a frontally-presented signal with this highly impoverished set of localisation cues.

The final experiment (Experiment 3b) demonstrated that despite the redundancy in the cues, *some* ITD information is crucial for this task. It seems that interaural level differences (ILDs) are ruled out as major cues for this task, as they were available but did not maintain performance in this experiment. This suggests that in Experiment 3a, subjects were reliant on ITDs in the envelope of the ongoing high-frequency signal. This is consistent with previous findings that envelope ITDs in high-frequency channels can be useful for sound lateralisation (Henning, 1974, 1980; McFadden and Pasanen, 1976). The surprising point is that this cue *alone* produced performance levels equal to those seen when robust low-frequency ITD information was available. This is unexpected, especially because several studies have demonstrated the high-frequency ITD to be a much weaker cue than the low-frequency ITD for lateralisation (Yost, 1976; Bernstein and Trahiotis, 1982).

An examination of Figure 3.9 reveals that for Experiment 3b, there were some effects at larger separations, with subjects reporting the perception of two sources despite the absence of ITD cues. It is possible that this is an artefact of the conflicting ITD and ILD cues at these extremes, where the ITD was zero but the overall ILD was non-zero as one of the source locations was displaced from the midline. Across subjects the mean overall ILDs at the extremes were measured to be 4 dB (test azimuth -21°) and -4.5 dB (test azimuth 21°). Indeed it was reported by subjects that some stimuli in this experiment produced an ‘unnatural’ percept, and this phenomenon has been reported previously for stimuli where interaural parameters are in conflict (e.g. Gaik, 1993).

## 3.8 General discussion

### 3.8.1 Subject performance

In an overview of this collection of data, it is apparent that there are strong individual differences between subjects. As discussed in section 3.4.4, efforts were made to

ensure that each subject behaved consistently across trials, but as expected the response criterion adopted by a particular subject was quite individualised. This individuality showed up mainly in the overall tendency of subjects to respond that they perceived two source locations in a stimulus presentation. In order to quantify this and inspect the individual differences, a 'false alarm' rate was calculated from the trials in which both noise sources emanated from the same location ('zero separation' trials). Across the entire testing set there were 250 of these trials, and for each subject the total percentage of false alarms was calculated. For S1, S2 and S3 this rate was very low (0, 6 and 1.5 % respectively) but for S4 the occurrence was higher (26%). This indicates that S4 had a greater tendency to respond positively to two source locations, i.e. a less strict response criterion. However, despite these inter-subject differences, that the overall patterns in the data were the same across subjects.

Furthermore, it is clear from the data that subjects could resolve the pair of noises with a high degree of certainty under the right conditions (determined by both spatial configuration and individual criterion). This is in contrast to a recent study by Braasch (2002), who presented similar stimuli to subjects in order to examine and model localisation in the presence of a distracter. Braasch presented concurrent 200 ms broadband noises separated in azimuth in the frontal region (analogous to our 0° reference location), and reported that in the majority of presentations, subjects perceived only a *single* auditory image. This is surprising since in that study the separation angles ranged from 15° to 90°, and in the present study subjects could fully resolve stimuli at the maximum separation of 21°. As the stimuli were identical apart from a small difference in their duration, the different responses of subjects are most likely related to the different tasks which they were asked to perform. Whereas our subjects were asked to report the number of sources they perceived, the primary task of Braasch's subjects was to give an estimate of the location of the auditory image (if there was only one) or the most lateral location (if there was more than one image). Perhaps the fact that our subjects were asked to focus exclusively on the number of sources allowed them to adopt a more sensitive criterion to this parameter.

It is important to consider exactly what cues the subjects in the present study may have used in their assessment of the number of sources in a stimulus. It is clear that the perception of more than one source was a binaural effect; noise pairs separated along the vertical midline remained perceptually fused. Taken together, the results suggested that the presence of two different ongoing ITDs in the presented

signal was responsible for separation. However, there are at least two possible ways in which this ITD cue might be used by the auditory system: (1) the two ITDs might be explicitly extracted from the composite signal, or (2) the presence of two ITDs in the mixed signal may be inferred from the neural representation. These two possibilities are discussed below.

### 3.8.2 A consideration of ITD sensitivity

If subject responses were driven by an ability to clearly perceive the two component ITDs, then it can be reasoned that the pattern of responses observed should bear some relation to ITD sensitivity with a single source. This idea is examined by analysing the results from Experiments 1, 2b and 2c (which spanned the five reference locations and were completed by all subjects) in terms of ITD. In Figure 3.10a, ITD thresholds are plotted as a function of reference ITD. Results for each of the four subjects are shown in a separate panel, and each panel contains data from Experiment 1 (broadband stimuli, red lines), Experiment 2b (high-pass filtered stimuli, blue lines) and Experiment 2c (stimuli with onset and offset ITDs removed, black lines). Reference ITDs and threshold ITDs were calculated from the impulse responses of the individual subjects as described previously (section 3.5.3) and the plotted values are an average of thresholds obtained in the vertical and horizontal conditions for each reference azimuth (ITD).

It can be seen that listeners in the present experiment required from 51 μs (S4) to 112 μs (S3) difference in ITD to consistently perceive two sources when they were presented concurrently around the frontal reference position (0° azimuth). Klumpp and Eady (1956) made measurements of ITD thresholds for listeners using headphones. They found that for a broadband noise with a reference ITD of 0 μs, the threshold ITD for determining whether a successive stimulus was to the left or right was 10 μs. This value is substantially smaller than that for the present experiment suggesting that, if ITD is the dominant cue in our task, it is more difficult to extract from a *concurrent* presentation of stimuli than from a sequential one.

According to Figure 3.10a, if an ITD difference drives resolution of these concurrent stimuli, then threshold ITD increases (in all conditions) with increasing reference ITD. A similar effect has been reported for sequentially presented stimuli

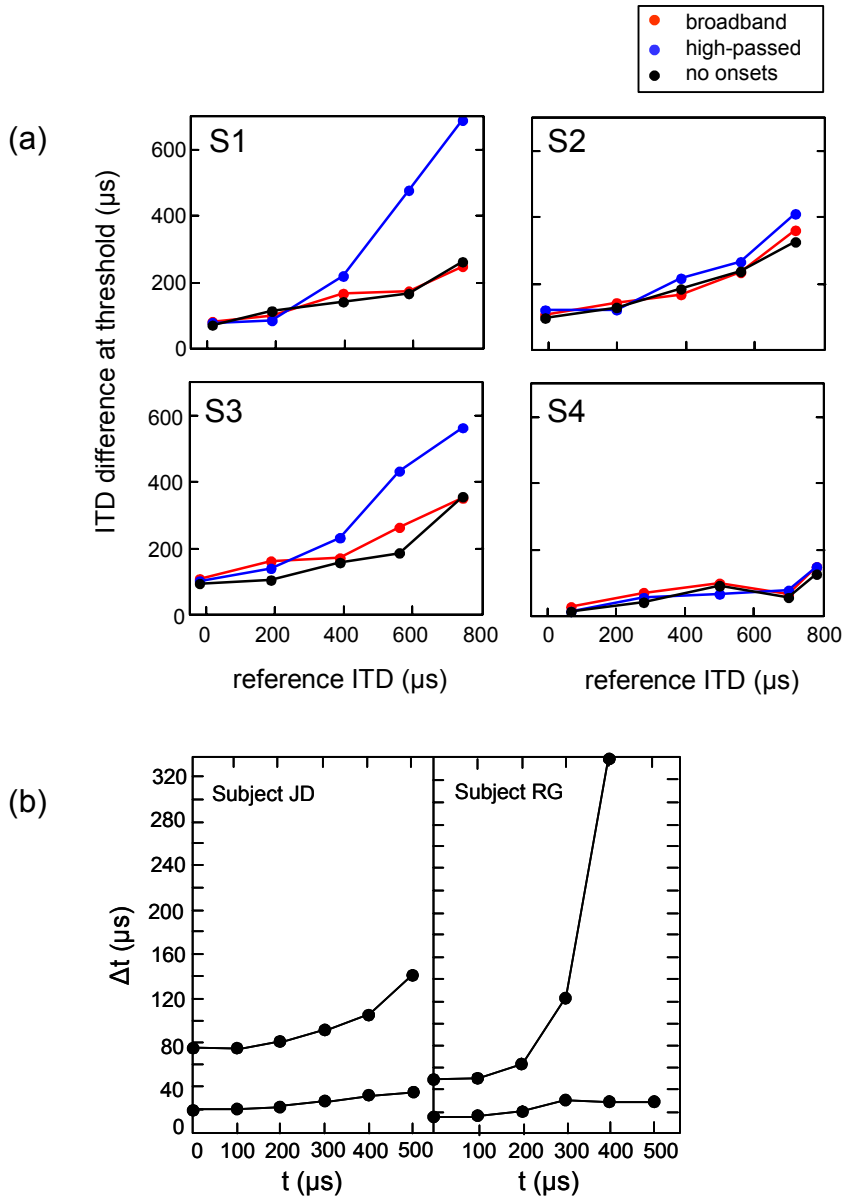



**Figure 3.10** (a) ITD thresholds as a function of reference ITD. The four subjects are represented in the four panels as labelled, and data is shown for broadband stimuli (Experiment 1, red lines), high-pass filtered stimuli (Experiment 2b, blue lines) and stimuli with onset/offset cues removed (Experiment 2c, black lines). (b) ITD thresholds as a function of interaural delay for two subjects using headphone presentation. Lower curves: low-frequency clicks (0.1-2 kHz). Upper curves: high-frequency clicks (3-4 kHz). Note that threshold increases with reference ITD, and that the gap between low and high frequency conditions increases with reference ITD. *Figure adapted from Hafter and De Maio, 1975.*

presented over headphones, where just-noticeable changes in ITD increase with baseline ITD. Klumpp and Eady (1956) used a band-pass filtered noise stimulus (150 – 1700 Hz) and found that the threshold for a 0 µs ITD stimulus was 9µs but increased to 29 µs and 50 µs for ITDs of 430µs and 790µs respectively. Other studies

have shown similar effects for high- and low-frequency transients (Hafter and Maio, 1975) and 500 Hz tones (Hershkowitz and Durlach, 1969; Domnitz and Colburn, 1977).

The data in Figure 3.10a also indicate that ITD thresholds are higher in general in the high-pass condition for the more lateral locations, confirming the observations made in Experiment 2b. This trend is most apparent for subjects S1 and S3. This finding is consistent with the study of Hafter and De Maio (1975) where it was found that just-noticeable ITD differences were larger for high-frequency (3-4 kHz) clicks compared to low-frequency (0.1-2 kHz) clicks. These authors also reported that this disparity increased with increasing baseline ITD, and for the largest ITD examined (500 μs) ITD discrimination performance using high-frequency clicks was unmeasurable in one subject. Their data (for two subjects) are shown in Figure 3.10b for comparison.

Thus, although there are substantial magnitude differences, it seems that there is some correspondence between the concurrent ITD thresholds estimated from the present data and reported sequential ITD discrimination thresholds. In both situations there is an increase in threshold with increasing reference ITD. In addition, whilst ITD in high-frequency regions is a useable cue, a low-frequency component improves performance at large values of ITD.

### 3.8.3 ITD extraction and interaural coherence

Although the analysis presented above is consistent with the idea that ITDs are useful cues for the separation of broadband noise sources, one must consider by what mechanisms the auditory system might recover these cues from the summed stimuli it receives. Preliminary computational modelling using biologically plausible elements indicates that a running cross-correlation of the time-domain inputs to the left and right ears may provide an adequate explanation of these data (e.g. Best *et al.*, 2002). The emergence of two peaks in the cross-correlation function suggests that the auditory system may be able to extract individual ITD estimates for each of the sources.

However, an important effect of two independent sources mixing is decorrelation of the signals at the two ears. It is possible that this decorrelation is

detected and used by subjects as an indication that there is more than one source. In terms of a cross-correlation mechanism, it may be that performance related less to the presence of two ITD peaks, and more to the *lack of a strong peak*.

It is certainly clear from the literature that listeners are sensitive to interaural correlation. Pollack and Trittipoe (1959a; 1959b) examined interaural cross-correlation JNDs for wideband noise sources. They found that JNDs were larger for a reference correlation of 0 as compared to 1, and also noted that low- and high-pass filtering increased JNDs. Gabriel and Colburn (1981) extended this work to look at bandwidth and level effects and reasoned that sensitivity to interaural correlation would define many basic binaural capabilities.

Of particular relevance to this study, it is known that interaural decorrelation is related to the perception of image width. Blauert and Lindemann (1986) mapped out the location and spread of images perceived in the head by their listeners when presented with headphone stimuli of varying interaural coherence. They found broadening and splitting as interaural correlation decreased from 1 to 0. This idea has also been applied in the free-field, and in an early study by Damaske (Damaske, 1967/68; described in Blauert, 1983, p. 246), listeners were presented with pairs of incoherent noises from loudspeakers, a situation almost identical to that of the present study. Damaske was interested in mapping out the auditory events perceived by the listeners under these conditions, and examples from two experienced listeners are shown in Figure 3.11. The top row shows two examples where distinct auditory events were perceived, and the bottom row shows that for certain configurations, a single broadened image is perceived. These results give some indication about the perceptual events experienced by listeners in the present study. It is likely that both broadening and splitting events were perceived, and individual subject criteria would determine at which point he/she would indicate hearing two source locations.

A consideration of the ideas presented above suggests that the perception of multiple sound sources under the conditions of the present study, while probably ultimately related to concurrent ITD estimates, must also be heavily influenced by interaural coherence. It is worth noting that decorrelation between the ears would not distinguish two sources from three, and an investigation of listeners' abilities to resolve three (or more) sources could help to define the contribution of different mechanisms. It is hoped that further experimentation and more sophisticated

modelling will illuminate in general the issue of how well the ITDs of concurrent stimuli are preserved in the auditory system.

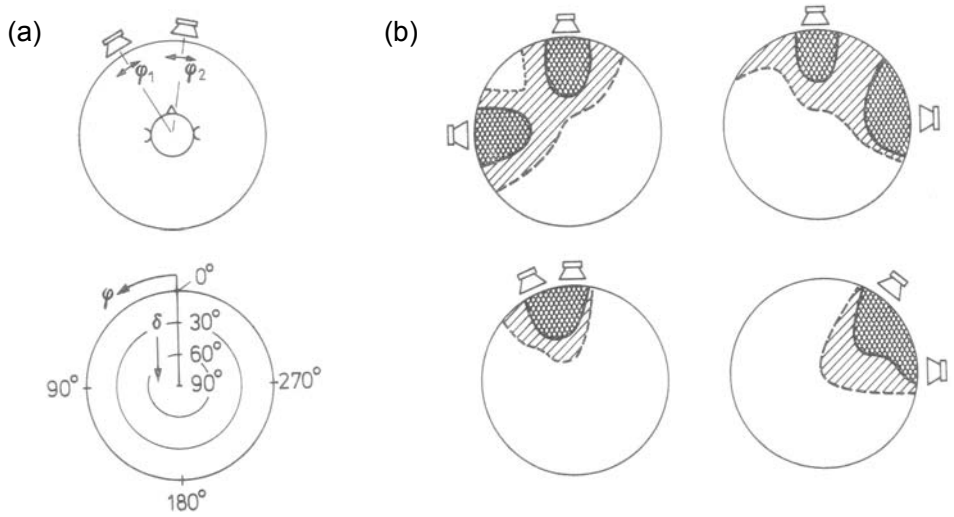



**Figure 3.11** Perceived auditory events under free-field stimulation with paired sources. Listeners were presented with pairs of incoherent noises from a distance of 2.6m. (a) Sources were placed at 18° elevation, but azimuthal locations could vary as indicated by the loudspeaker icons (upper panel). Perceived auditory events were mapped onto an equal-area projection of the upper hemisphere (lower panel). (b) The two levels of shading represent ranges within which auditory events were perceived with different frequencies. The top row shows two examples where distinct auditory events were perceived, and the bottom row shows two examples where a single broadened image was perceived. *Figure from Damaske, 1967/68, described in Blauert, 1983, p.246.*

### 3.8.4 ITD as a cue in other situations

It is important to note that the data in this study were collected under anechoic conditions, and the acoustics of real-world listening conditions are very likely to affect the ability of subjects to discriminate one source from two. Certainly the presence of echoes could be expected to confuse the estimate of source numbers (by increasing the number of peaks and/or the interaural decorrelation). On the other hand, it has been shown that the impact of reverberation on the localisation of a single source is minimal in some cases (Hartmann, 1983; Shinn-Cunningham, 2000).

Furthermore, the strict reliance on ITD differences for resolving concurrent sources observed in this study, and the robustness of the ITD cue, is somewhat of a

contrast to previous studies that have shown ITD to be of limited use in certain competing speech tasks. For example, when presented with simultaneous speech sounds, listeners cannot group components across frequencies on the basis of ITD alone (Culling and Summerfield, 1995). Furthermore, Hukin and Darwin (1995) reported that a single harmonic cannot be segregated from a vowel by just altering its ITD. This is likely to be a result of competition between location cues and other strong grouping and streaming cues (such as frequency content and continuity) in these stimulus paradigms; presumably the auditory system weights the available cues differently depending on the nature of the stimuli in order to best represent the auditory scene.

### 3.8.5 Ineffectiveness of spectral cues in a mixed signal

It was interesting, although not altogether surprising, to find that pairs of broadband noises could not be resolved when separated along the median vertical plane (with no binaural differences present). This indicates that the location-dependent spectral cues of the individual sources were not accessible in the mixed signal under these conditions.

Of relevance to this idea, Rakerd and colleagues (1999) examined the interaction of identification and localisation in the median plane. Subjects were familiarised with different broadband stimuli containing dips or peaks (these were named A, B, etc.). A stimulus chosen randomly from this set was then played from a location on the median vertical plane, and subjects had to identify the stimulus. It was found that roving the location greatly impaired performance if relevant identity features were at high frequencies. This suggests that given no other information, the auditory system could not disentangle the source spectrum and the location-related spectrum from the available combined spectrum. In the present study, although the source spectrum was known to be flat, the two different spectra corresponding to the two different stimulus locations could not be disentangled from the composite signal. This problem may not hold when stimuli do not fully overlap in time and/or frequency, and this idea is explored in the chapters to follow.

## 3.9 Conclusions

Spatial resolution with concurrent sources was examined systematically with a focus on the role of the different localisation cues. The study was unique in that it (a) employed broadband stimuli that were distinguishable *only* on the basis of location, (b) focussed on the detection of multiple sources in an acoustic stimulus rather than on issues of identification and/or localisation, and (c) examined stimuli distributed both horizontally *and* vertically. It was shown that a concurrent pair of broadband white noises could be resolved if a sufficiently large binaural difference was present. In particular, differences in ongoing ITDs were shown to be robust cues for separation, including those in the envelopes of high-frequency channels (above 2 kHz).

# Chapter 4: Sound localisation in the presence of a concurrent masker

## 4.1 Introduction

In the previous chapter, spatial resolution was examined in a most extreme case, where the stimuli to be resolved spatially were indistinguishable. In a more realistic situation, it is unlikely that competing sounds would be identical. In regular environments, different auditory objects have different structures and identities, although many will overlap in frequency and in time to a large extent. This chapter will focus then on broadband stimulus pairs that have different temporal structures that make them distinguishable from each other. Furthermore, given a set of unique auditory objects, the situation that will be the focus of this chapter is when a listener is required to make a spatial judgement about only one of the sources making up the competing pair. This situation is again more relevant to everyday experience where one is generally interested in the whereabouts and content of a particular source of interest in the presence of interfering sources.

The issue of how well a target auditory object can be localised in the presence of other objects is one that has received surprisingly little attention in the literature. If we appreciate the complexity of the processing underlying spatial hearing (see Chapter 1), then the situation where there is interference from a competing source is extremely interesting. There are at least three forms of interference that can be imagined in this situation: (a) acoustic interference at the most peripheral stages that may hinder the reception of the target signal, (b) analytic interference at the various stages where the acoustic cues to location are extracted and processed, and (c) perceptual interference, where the listener's ability to map a target object and attend to it may be obstructed by the presence of other objects.

## 4.2 Previous studies of concurrent localisation

The studies that have attempted to investigate sound source localisation in the presence of a masker have produced a range of results. Some of these results appear inconsistent with each other, and the conclusions drawn from them quite different. A striking example is that many researchers have reported that target stimuli tend to be repelled away from the location of a masker when it is present (referred to here as a 'pushing effect'). On the other hand, several reports have described quite the opposite, where the target is mislocalised with a shift *towards* the masker ('pulling effect').

### 4.2.1 Pushing effects

In the horizontal plane, several authors have reported pushing effects of maskers on target sound localisation using stimuli presented from loudspeakers in an anechoic room. Suzuki and colleagues (Suzuki *et al.*, 1993; Suzuki and Sone, 1986) examined the localisation of a 1 sec target stimulus embedded in a 2 sec pink masking noise. They found that the target was generally pushed away from the noise, and this was true for pure tones, narrow-band noises, and even a pink noise identical to the masker. Canévet and Meunier (1996) used a frontal 4 kHz tone masker and examined horizontal localisation of a 50 ms 4 kHz test tone. The masker preceded the target by a variable amount of time (20 ms to 10 sec). For the shortest onset time difference (20 ms), the authors saw precedence-type effects, with perception being dominated by the leading frontal source. For onset differences of 100 ms and greater however, they saw a reversal in this effect, with the target being pushed away from the masker position. In a recent study, Braasch and Hartung (2002) presented two broadband sounds in individualised virtual auditory space, a 500 ms masker with an embedded 200 ms target. With a signal-to-noise ratio of 0 dB, a pushing effect was seen for masker locations of -90°, 0° and 90° azimuth.

Only one author has reported similar effects in the vertical dimension. Getzmann (2002; 2003b) examined the influence of background noise on vertical localisation in the median plane. He presented 500 ms pink noise targets against a 2 sec background pink noise (with a 1 sec onset delay) at a favourable signal-to-noise

ratio of 5 dB. It was found that targets were shifted away from the distracter by about 3-5° on average.

### 4.2.2 Pulling effects

In contrast to the above reports, several similar experiments conducted in the horizontal plane have produced the result that a concurrent masking sound exerts a pulling effect on target localisation. In an early example (Butler and Naunton, 1964), it was found that the perceived location of random noise pulses (at various horizontal locations) in the presence of continuous interference (random noise at one ear) was shifted towards the ear containing the interference. This effect was larger as signal duration was increased (16° at 0.5 ms to 30° at 100 ms). Thurlow *et al.* (1965) presented a short pulse (0.1 ms) and a longer 1500 Hz tone with simultaneous onsets. Subjects localised the tone, and it was found that the pulse had a pulling effect, but only for relatively short duration tones (40 to 400 ms) and not for longer continuous tones (up to 2 sec). In a more recent study, Heller and Trahiotis (1996) looked at lateralisation by ITD of a high-frequency sinusoidally-amplitude-modulated (SAM) tone in the presence of a simultaneous lower-frequency SAM tone (same 250 Hz modulation rate). They reported that targets were pulled towards the masker, and put this forward as evidence that concurrent sounds source interference is not restricted to within-frequency channel effects.

### 4.2.3 Other studies

There are other studies that have examined the effect of a masking sound on target source localisation and have not found consistent effects like those described above. Lorenzi *et al*. (1999) examined the horizontal localisation of 300 ms click trains embedded in a 900 ms noise burst. For signal-to-noise ratios greater than 0 dB, these authors saw no effect on localisation, and for poorer signal-to-noise ratios they saw both pushing and pulling effects, dependent on the individual listener.

Only a few studies examined concurrent sound localisation for stimuli located all around the listener (most have concentrated on the horizontal plane only). Good and colleagues published two studies (Good and Gilkey, 1996; Good *et al.*, 1997) in

which subjects localised click trains (269 ms) that were temporally centred in noise (468 ms). The location of both the clicks and the noise were varied on a sphere surrounding the listeners, and subjects responded by pointing at a 20 cm diameter spherical model of auditory space. They examined the effect of the masker on three different components of localisation estimates: left/right, front/back and up/down. It was found that performance was increasingly disrupted for increasingly poor signal-to-noise ratios. The dimension in which the masker was extreme was most affected, e.g., when the masker was to the left or the right, localisation in the left/right dimension was most affected. They did note some tendency to bias the target towards the masker at poor signal-to-noise ratios.

In their elegant study, Langendijk *et al*. (2001) used individualised virtual auditory space to examine the localisation of a train of noise bursts in the presence of one or two distracters (a harmonic complex and/or a frequency-swept complex tone). All three signals were well-localised when presented on their own. In contrast to previous studies, the 'distracters' were so-named because they were played in the silent gaps between the target noise bursts such that the different signals did not overlap in time (this approach was chosen by the authors to reduce classical 'masking' effects). Targets were presented from one of 85 positions on the sphere and distracters from 17 positions. It was found that localisation errors increased in the presence of a distracter (more so in the presence of two) and that the greatest impact occurred when the horizontal distance between the target and the distracters was small. Unfortunately the authors did not report on the direction of these errors (i.e. pushing or pulling) although they did report a bimodal distribution of responses in about 8% of their one-distracter trials, indicating 'attraction to' or 'confusion with' the distracter.

### 4.2.4 A reasonable hypothesis?

The large majority of the studies of concurrent sound source localisation discussed above have adopted stimulus presentation paradigms where the target and masker do not have simultaneous onsets (Suzuki and Sone, 1986; Suzuki *et al.*, 1993; Canévet and Meunier, 1996; Getzmann, 2002, 2003b; Braasch and Hartung, 2002; Lorenzi *et al.*, 1999; Good and Gilkey, 1996). In general, the masker onset precedes the target onset by some amount (from 20 ms up to as long as 10 sec) presumably to increase

the salience of the target when it is sounded, and to ensure that it is clearly distinguished from the masker. It has long been known that simultaneity enhances the grouping of objects in auditory scenes, with the rationale that sounds from a common object will usually coincide in time (this idea is explored in depth in Bregman, 1990).

A review of the previous sections reveals that this stimulus arrangement is most commonly found in studies where 'pushing effects' of the masker were observed. In contrast, the studies reporting 'pulling effects' (Butler and Naunton, 1964; Thurlow *et al.*, 1965; Heller and Trahiotis, 1996) presented the target and masker with simultaneous onsets. In one of the only studies that examined both situations, Canévet and Meunier (1996) attributed their observed pushing effects to adaptation to a leading masker, and in fact showed that the lead-time of the masker influenced the size of the shift. They also reported that when the two tones were presented simultaneously, a fused single percept resulted. It seems reasonable then to assume that simultaneity contributes to this complex set of localisation phenomena. Perhaps the pushing effects are an artefact of the auditory system streaming the concurrent pair on the basis of onset differences, and the pulling effects occur when the concurrent objects are grouped. This distinction may reflect 'synthetic' and 'analytic' listening modes as described by Dye and colleagues (Dye, 1990; Dye *et al.*, 1994), where binaural information is either combined across frequencies ('synthetic') or processed separately for different perceptual objects ('analytic').

## 4.3 Approach

Two experiments were carried out to examine further the spatial perception of concurrent sound sources. In order to investigate localisation effects in detail, broadband stimuli were chosen that were well-localised in all dimensions. In order to identify a target and a masker, however, the stimuli were made distinguishable on the basis of their envelope. The basic arrangement was to use a train of noise bursts as the target, and a continuous noise as the masker. Locations in this set of experiments were confined to the frontal hemisphere of space in order to concentrate our analysis on the left-right and up/down dimensions without considering in detail the front-back dimension. The experiments (and the entire thesis) are primarily concerned with

simultaneous stimuli, and thus we create a situation where the signals to be processed by the auditory system overlapped significantly in *both frequency and time.*

The aim of Experiment 1 was to test the hypothesis that simultaneous target and masker onsets result in a pulling of perceived location, but that non-simultaneous onsets result in a pushing effect. In order to probe this idea, localisation of a target in the presence of a masker was compared for simultaneous stimuli, partly-simultaneous stimuli (masker onset precedes target onset), and completely sequential stimuli. For simultaneous stimuli, the effect of stimulus duration was also examined, as previous studies have reported a change in the magnitude of observed effects with changes in exposure time (e.g. Butler and Naunton, 1964).

Experiment 2 was not concerned with temporal aspects, but with stimulus content. Completely simultaneous stimuli were employed, and two issues that arose out of Experiment 1 were probed. The first two conditions examined the impact of target dominance by varying the relative strength of target and masker. In a third condition, the involvement of low-frequency spatial cues was examined by removing low-frequency stimulus content.

## 4.4 Experimental methods

### 4.4.1 Subjects and task

Four subjects (S1, S3, S5, S6) participated in the experiments. Subjects S1, S3 and S5 were highly practiced in auditory localisation experiments, whereas subject S6 had less experience. Experiments were carried out in the anechoic chamber and stimuli were presented in virtual auditory space (Chapter 2).

The task was an absolute localisation task, and the response paradigm was the same as that described in section 2.4.1. Briefly, subjects were positioned in the centre of the chamber and were required to point their nose to the perceived target location on each trial. An electromagnetic head-tracker recorded their estimates after a response button was pressed. No specific training was carried out for this experiment, as all subjects had been trained to localise in a similar set-up previously (section 2.4.2).

The two experiments each consisted of a number of conditions (five and three for Experiments 1 and 2). Details of the stimulus manipulations characterising each condition are given in the individual sections to follow. Experiment 1 was completed before the commencement of Experiment 2. In total, each subject completed 40 tests (approximately 7 hours of listening time). Testing was completed over a period of approximately eight weeks.

### 4.4.2 Stimulus configurations

In order to generate target/masker stimulus pairs, the two signals were filtered with DTFs corresponding to the two desired locations and then added together. After summing, stimuli produced a sensation level of approximately 50 dB.

All stimuli were presented in the frontal hemisphere of space. Stimulus locations are described in this chapter using lateral and polar angle co-ordinates (see section 2.2). Three masker locations were chosen on the horizontal plane passing through the ears: (-30°,0°), (0°,0°) and (30°,0°) (Figure 4.1a). In each target/masker trial, the masker was presented from one of these three locations and the target from one of 19 locations surrounding the masker location. One of the 19 target locations coincided with the masker location, six differed in lateral angle only (i.e. shared the same polar angle), four differed in polar angle only (i.e. shared the same lateral angle) and the remaining eight differed in both lateral and target angle (see Figure 4.1b). The minimum spacing of target and masker (aside from the coincident configuration) was 15° and the maximum spacing was 60°.

In a single test, each of the 57 target/masker combinations was presented once in a random order (3 masker locations x 19 target locations). Interleaved with these were 35 control trials where the target was presented in isolation at each of the unique target locations. Thus a test comprised 92 trials in total, and took around 10 minutes to complete. Each subject completed five such tests to obtain five estimates of the perceived location of all stimuli. These five localisation tests were completed in succession, but the different condition blocks were undertaken in a random order by each subject.

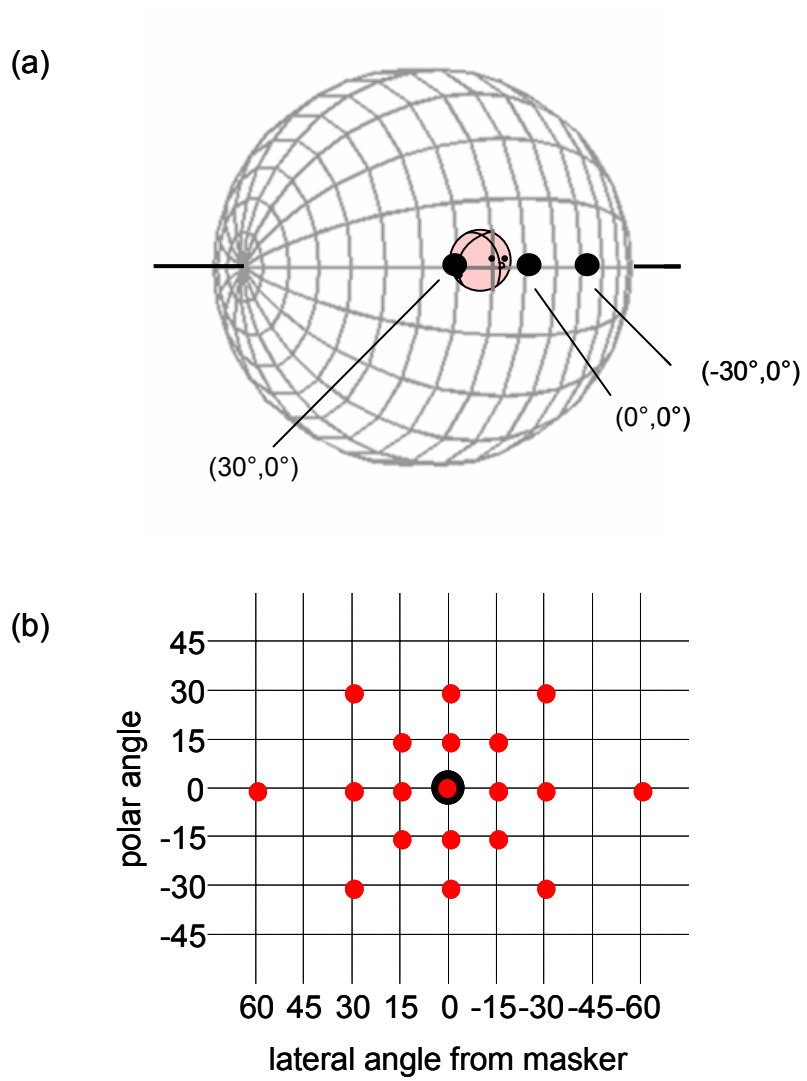



**Figure 4.1** Stimulus configurations, described using the lateral/polar co-ordinate system. (a) The masker could occupy one of three locations as indicated by the black dots: (0°, 0°) (directly in front), (-30°, 0°), and (30°, 0°). (b) For each masker location (black dot), 19 target locations (red dots) were distributed in both lateral and polar angle as depicted.

### 4.4.3 Data analysis

The spherical correlation coefficient (SCC, see section 2.4.3) was used as a global measure of localisation performance. For the control trials, of which there were 175 for each subject per condition, the SCC was useful for measuring a subject's baseline localisation accuracy. For trials in which the masker was present (285 trials per subject per condition), the SCC was calculated and compared to the control SCC in order to gauge the overall impact of the masker on localisation performance. As the

number of estimates going into the SCC calculation can affect the outcome of the calculation, this measure was taken only as an indicator of overall performance in control and test conditions. Detailed comparisons between conditions were avoided because of the different population sizes.

Once the overall performance levels were established, it was useful to inspect the data for specific effects of the masker on localisation. As there were five replicate responses obtained for each stimulus configuration, these were pooled and centroids were calculated. Thus for each target location, the centroid under no-masker and masker conditions could be compared. These centroid pairs, and the relevant masker locations, were plotted on spheres representing the sphere of space for visualisation.

In order to quantify the influence of the masker on binaural and non-binaural localisation, centroids under masker and no-masker conditions were decomposed into lateral and polar angle components. By plotting actual lateral (or polar) angle against perceived lateral (or polar) angle it was possible to observe any systematic effects in these two dimensions.

## 4.5 Experiment 1: Influence of simultaneity and duration

### 4.5.1 Stimuli

There were five stimulus conditions (depicted in Figure 4.2). In all conditions, the target stimulus was a train of 30 ms noise bursts with 30 ms silent intervals (16 Hz cycle). Each noise burst was ramped on and off with a raised cosine of 15 ms duration (i.e. continuous ramping) and each burst was independently generated. The masker was a continuous noise burst with 15 ms raised cosine ramps at each end. Both stimuli had a frequency range of 300 Hz to 16 kHz. Stimulus levels were chosen to create a pair of stimuli (target and masker) of approximately equal loudness. This was achieved by matching the average overall RMS level of the two stimuli (before DTF filtering and adding). Note that this matching produces an overall ‘target-to-masker ratio’ of 0 dB.

In condition A (‘concurrent 270’), the target and masker had simultaneous onsets and offsets and the target consisted of five noise bursts, giving a duration of 270 ms. Condition B (‘lead 270’) was identical except the masker began 270 ms

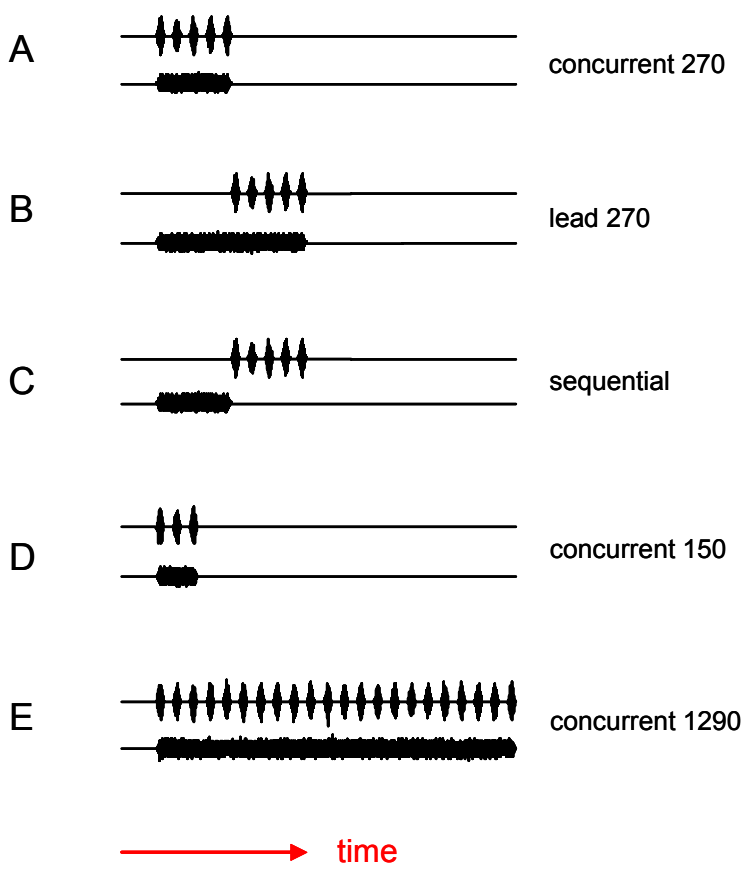



**Figure 4.2** The five stimulus conditions of Experiment 1. The target was a train of noise bursts and the masker was a continuous noise in all conditions. Condition A ('concurrent 270'): target and masker were completely simultaneous with a duration of 270 ms. Condition B ('lead 270'): the masker was 540 ms long and preceded the target by 270 ms. Condition C ('sequential'): the masker and target were both 270 ms long but the masker fully preceded the target. Condition D ('concurrent 150'): target and masker were completely simultaneous with a duration of 150 ms. Condition E ('concurrent 1290'): target and masker were completely simultaneous with a duration of 1290 ms.

before the target (and hence had a total duration of 540 ms). Condition C ('sequential') was designed to remove concurrency: the masker was sounded for 270 ms and then silenced as the target was played for 270 ms immediately following. Conditions D and E again consisted of simultaneous stimuli, like condition A, but the total duration of the pair was varied. In condition D ('concurrent 150'), only three noise bursts made up the target train giving a total duration of 150 ms. In condition E ('concurrent 1290'), the target consisted of 22 noise bursts giving a total duration of 1290 ms.

## 4.5.2 Results

### Overall performance

All subjects localised to a good degree of accuracy in the control (no masker) trials of each condition. Table 4.1 summarises the performance in the control trials for each subject under each stimulus condition. All subjects localised the single noise-train well, with the SCC ranging from 0.83 to 0.95. Overall, subjects S3 and S5 were more accurate localisers than subjects S1 and S6.

When presented in conjunction with the masker, target localisation was affected, but to a surprisingly minor extent overall. Table 4.1 also summarises the performance in the masker trials for each subject under each stimulus condition. The SCC ranges from 0.63 to 0.96, and values in the majority of conditions are not strikingly different from the corresponding control sets. In general, it appears that subjects S3 and S5 were less affected by the presence of the masker than were S1 and S6. Interestingly these subjects were also the better localisers, suggesting that accuracy and robustness of localisation may co-vary. Finally, it appears from this metric that S1 and S6 were particularly disturbed by the masker in the 'concurrent 270' condition.

**Table 4.1** Spherical correlation coefficients (SCCs) for each of the four subjects (S1, S3, S5, S6) in control (no masker) trials and test (with masker) trials of Experiment 1. Each of the five rows contains values for the five stimulus conditions. Each SCC was calculated from 175 control trials or 285 test trials. See section 2.4.3 for details of this statistic.

| | | S1 | S3 | S5 | S6 |
|---|---|---|---|---|---|
| concurrent 270 | cont | 0.84 | 0.92 | 0.93 | 0.85 |
| | test | 0.70 | 0.92 | 0.90 | 0.63 |
| lead 270 | cont | 0.88 | 0.94 | 0.94 | 0.88 |
| | test | 0.73 | 0.94 | 0.92 | 0.81 |
| sequential | cont | 0.86 | 0.95 | 0.95 | 0.90 |
| | test | 0.89 | 0.96 | 0.95 | 0.88 |
| concurrent 150 | cont | 0.88 | 0.94 | 0.93 | 0.88 |
| | test | 0.88 | 0.94 | 0.90 | 0.75 |
| concurrent 1290 | cont | 0.83 | 0.94 | 0.94 | 0.88 |
| | test | 0.82 | 0.95 | 0.93 | 0.82 |

Qualitative analysis of responses

It was clear that the localisation effects induced by the masker, although not extreme, varied between individuals and showed a different pattern in the different conditions. Figures 4.3-4.7 contain spherical plots of the results from the five conditions. In each figure, there are four rows depicting results from the four subjects. Each row contains three spherical plots, each showing data for one of the three masker locations (left: 30°,0°; middle: 0°,0°; right -30°,0°).

On each spherical plot, the masker location is depicted by a large red dot and the location directly in front of the listener is indicated by the grey circle or 'nose'. Shown on these spherical plots are the centroids of localisation estimates in the presence of the masker (blue dots) in comparison to the centroids of localisation estimates in the control condition (red dots) with a blue line joining corresponding pairs. The influence of the masker is briefly described here, with 'push' and 'pull' used to define apparent shifts in target localisation away from or towards the position of the masker.

In the 'concurrent 270' condition (Figure 4.3), the masker produced disruptions in elevation estimates. The influence varied between subjects: a small push in S3, a noticeable pull in S6, and an inconsistent effect in the other subjects. For all subjects, the masker appeared to induce a small but consistent bias in lateral angle estimates of the target away from the masker.

In the 'lead 270' condition (Figure 4.4), a similar pattern of errors was observed. Elevation errors were more often observed to be a shift away from the masker, and lateral angles again showed a consistent shift away.

The 'sequential' condition was quite different in that the masker had very little effect on localisation (Figure 4.5); the control and test estimates are generally in very close agreement for all subjects. Some minor elevation errors were displayed by S1 and S6 (who are poorer localisers in general).

In the short and long duration conditions ('concurrent 150' and 'concurrent 1290'; Figures 4.6 and 4.7), the general pattern of errors was not too dissimilar from that of the medium duration condition ('concurrent 270'). Some minor disruptions of elevation were observed, especially a pulling effect in S6 and a pushing effect in S3 and S5 under the long duration condition. In both conditions, subjects showed a lateral angle bias away from the location of the masker. The only exception to this

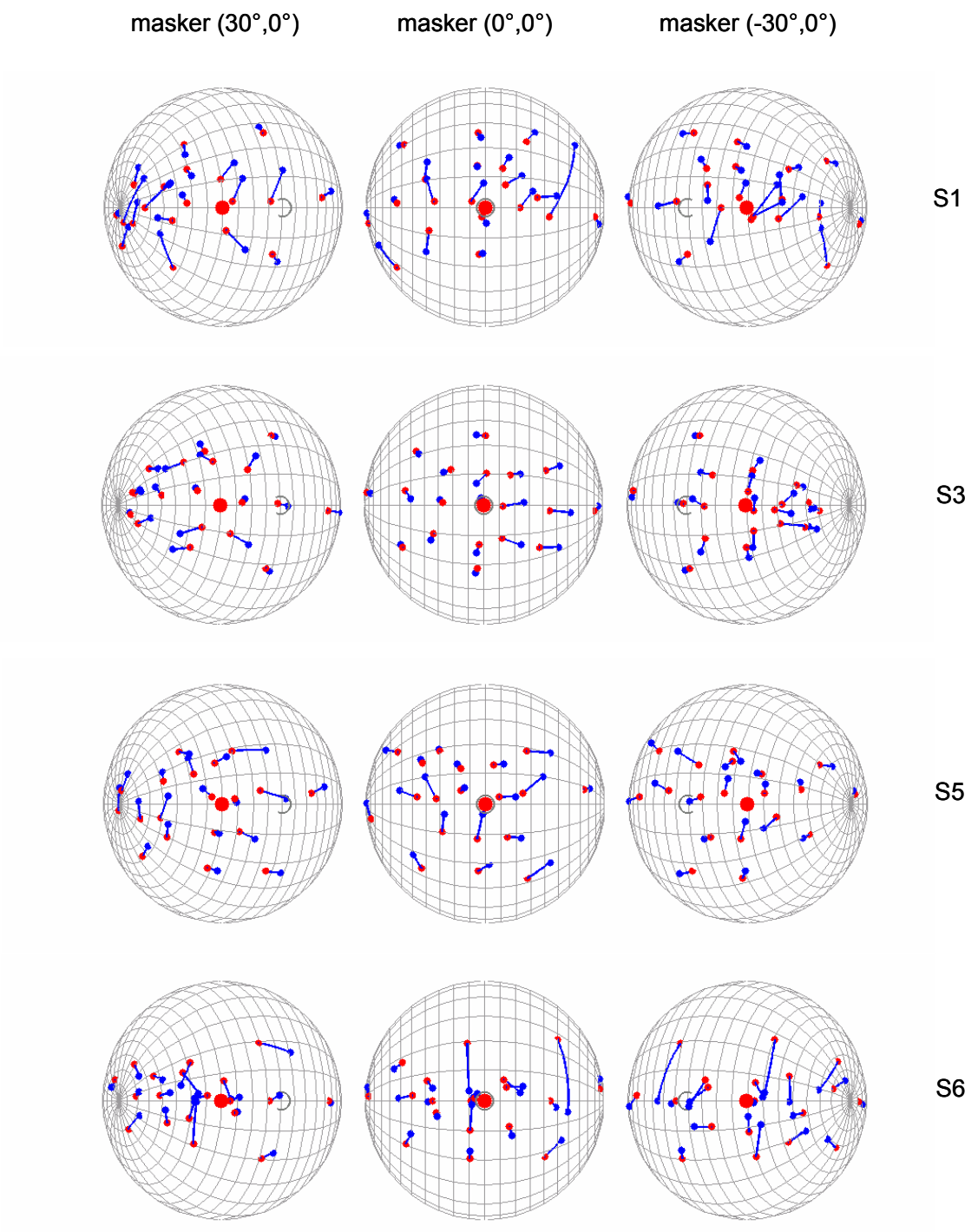



**Figure 4.3** Spherical plots for the 'concurrent 270' condition of Experiment 1. The four rows depict results from the four subjects, and the three spherical plots in a row show data for the three masker locations (left: 30°,0°; middle: 0°,0°; right -30°,0°). On each plot the masker location is depicted by a large red dot and the location directly in front of the listener is indicated by the grey circle or 'nose'. Red dots show the centroids of localisation estimates in the control condition. Joined to these by blue lines are blue dots representing the centroids of corresponding estimates in the presence of the masker.

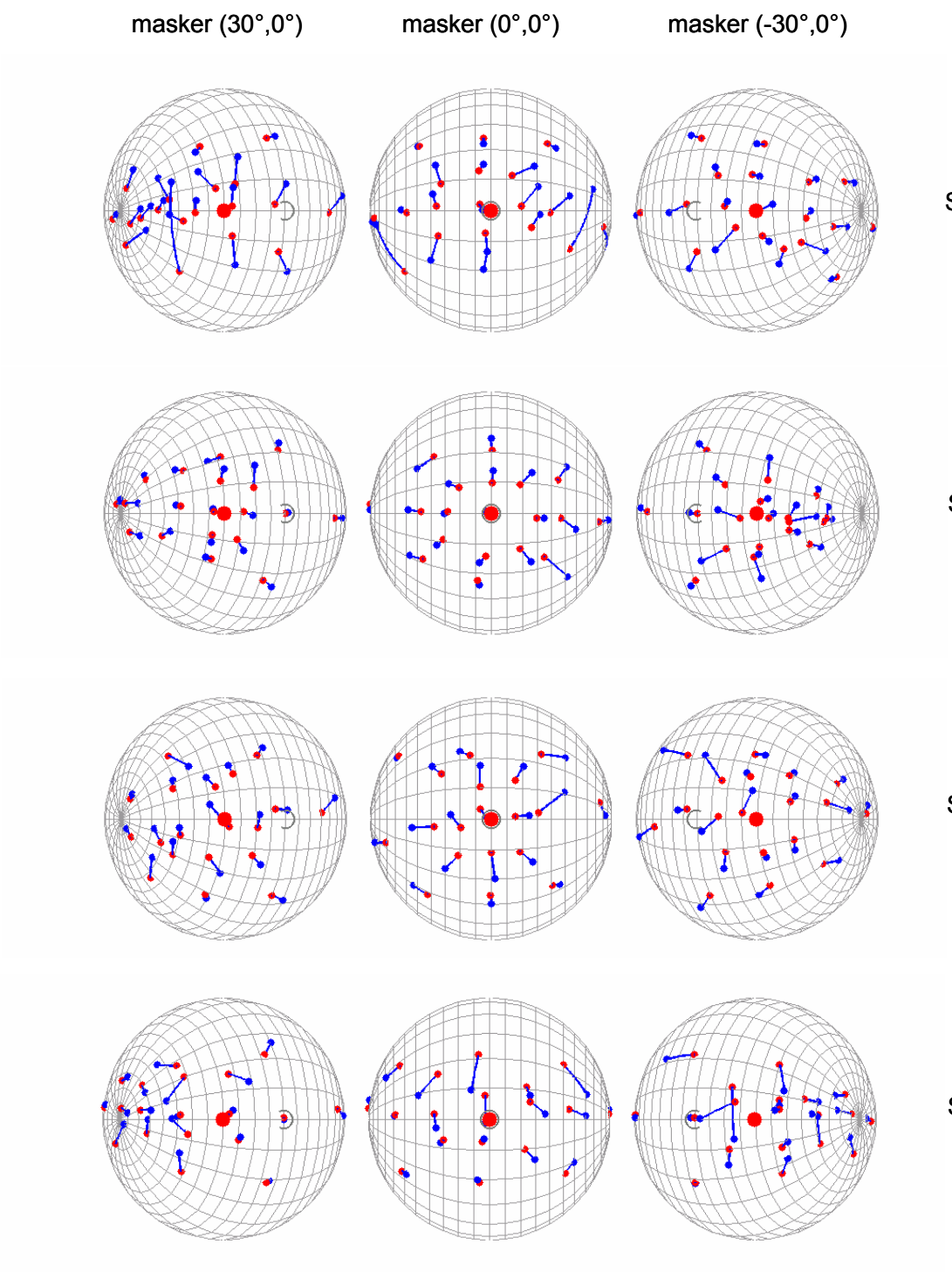



**Figure 4.4** Spherical plots for the 'lead 270' condition of Experiment 1. All other details as for Figure 4.3.

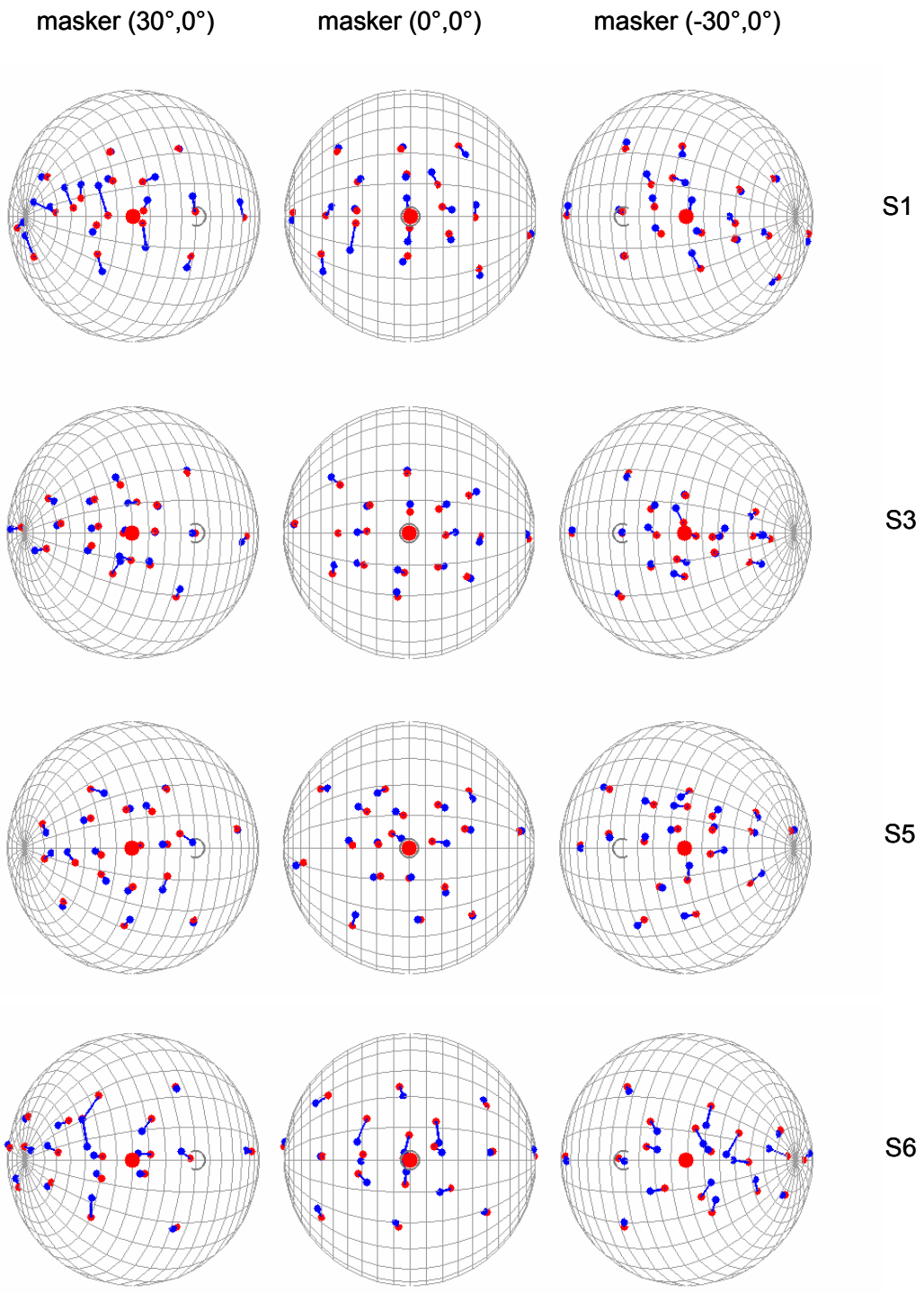



**Figure 4.5** Spherical plots for the 'sequential' condition of Experiment 1. All other details as for Figure 4.3.

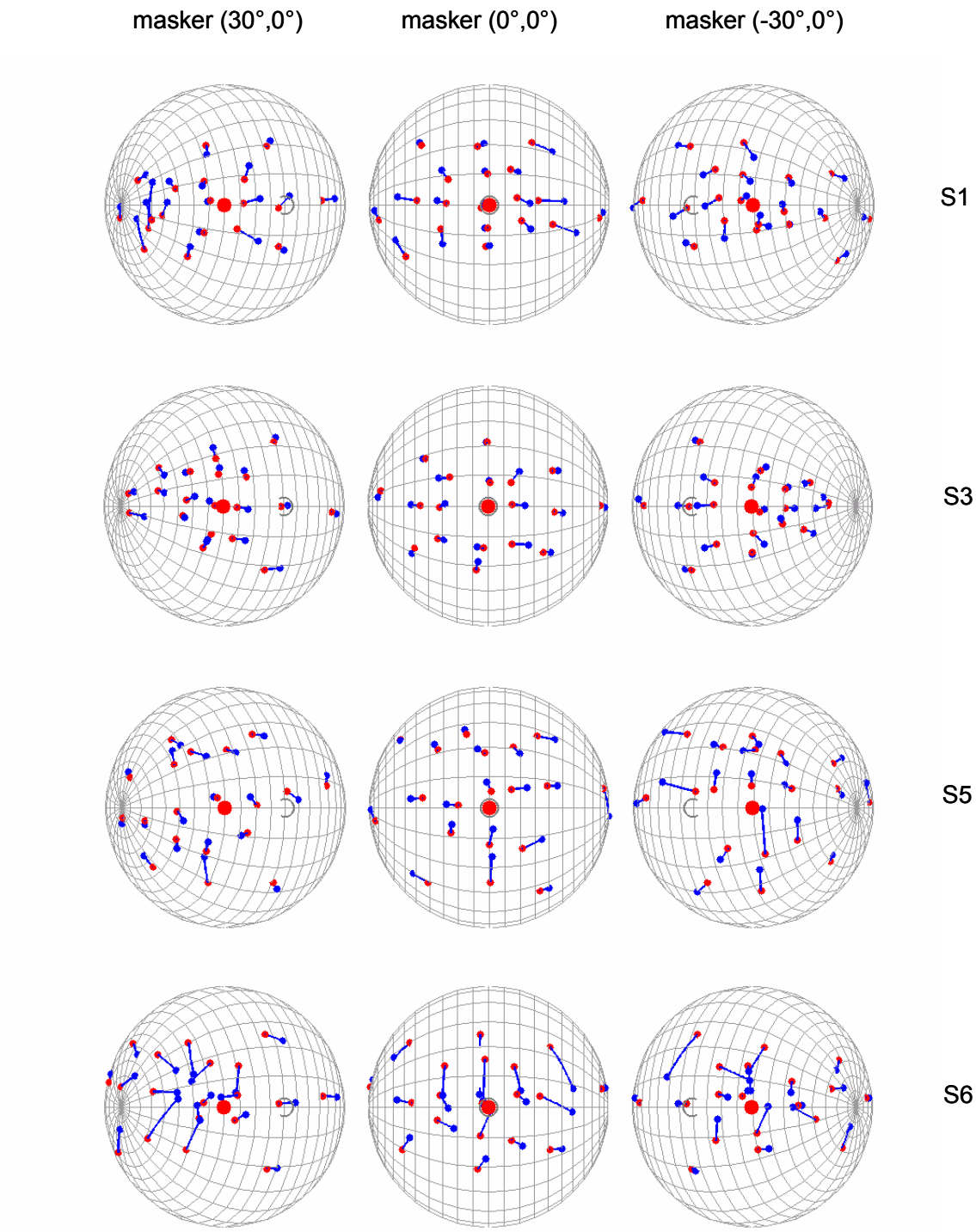



**Figure 4.6** Spherical plots for the 'concurrent 150' condition of Experiment 1. All other details as for Figure 4.3.

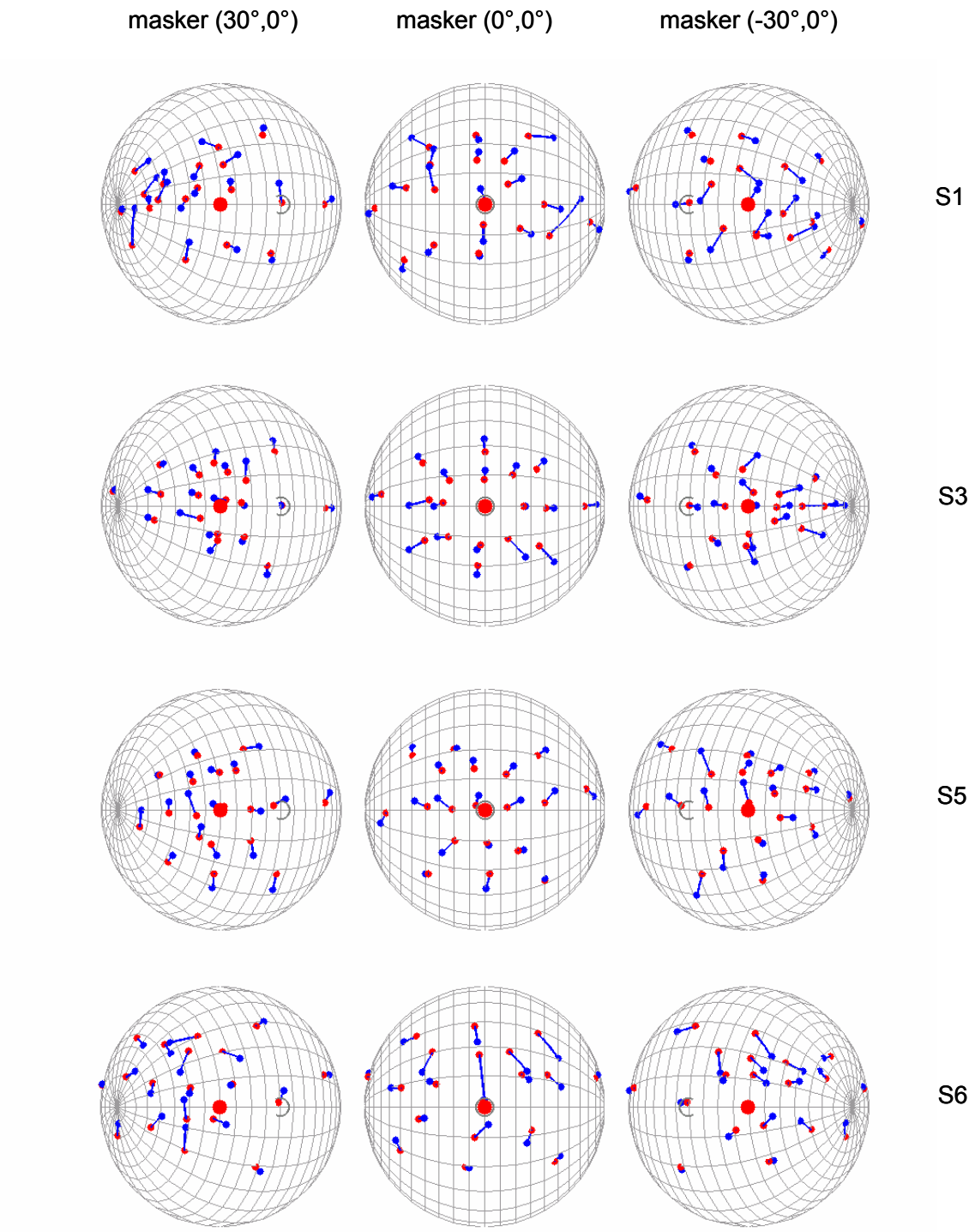



**Figure 4.7** Spherical plots for the 'concurrent 1290' condition of Experiment 1. All other details as for Figure 4.3.

general trend was S6 who showed some bias *toward* the masker for proximal locations (this appears to be due to confusion or fusion of the target and masker).

### Quantification of localisation bias

In order to investigate the size and nature of the biases caused by the masker in each condition, the difference between masker and no-masker centroids was calculated and expressed as lateral and polar angle shifts. These shifts were expressed relative to the masker lateral or polar angle, such that a positive shift indicates a bias away from the masker and a negative shift indicates a bias towards the masker. Data from all masker locations was pooled, but individual subjects were treated separately in order to capture the variability seen in the spherical plots.

Figure 4.8 illustrates the mean lateral angle shifts. The five clusters of bars represent the five conditions, and the individual bars represent the individual subjects. Note that all mean lateral angle shifts (except one) are positive, indicating a bias away from the masker lateral angle. Furthermore, a previous observation is confirmed, in that the sequential condition showed far less effect than the other conditions. The error bars show the standard error of the mean (SEM) and the asterisks indicate shifts that were significantly different from zero (two-tailed t test, $p<0.05$). In the concurrent conditions, all subjects showed a positive shift, and this was significant in all cases but one (S6, 'concurrent 150'). In the sequential condition, two subjects showed a significant positive shift, but the shifts were much smaller than in the other conditions.

An ANOVA on the group data revealed a significant effect of condition on the lateral angle shift [$F(4,1135) = 10.1$, $p < 0.001$]. Post-hoc analysis (Tukey HSD, $p=0.05$) showed that shifts were significantly smaller in the 'sequential' condition than in all of the concurrent conditions. There was no significant difference in the lateral angle bias between the 'concurrent 270'and 'lead 270' conditions, suggesting that the leading portion of the masker did not influence this effect. Furthermore, varying the duration of a concurrent stimulus presentation did not produce significantly different effects in the binaural dimension.

Figure 4.9 illustrates the mean polar angle shifts. Again, each cluster of bars represents one of the five conditions, the individual bars represent the individual subjects, the error bars show SEMs, and the asterisks indicate significant shifts. The individual differences are much more striking than in the lateral angle analysis. The shifts are distributed on either side of zero, which is consistent with what was seen in

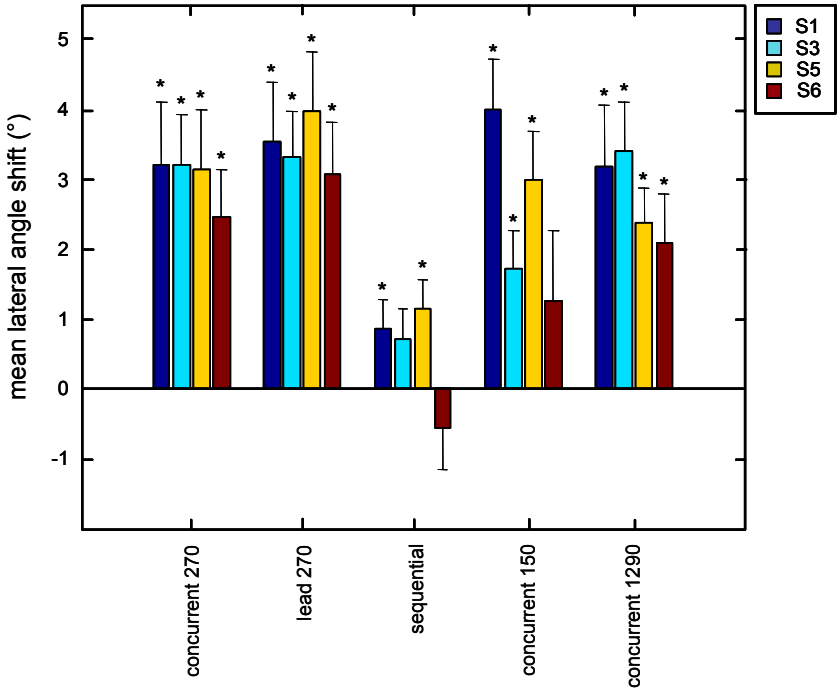



**Figure 4.8** Mean lateral angle shifts (across all locations and maskers) in Experiment 1. The five groups of bars show the five stimulus conditions and each bar in a group represents a different individual. A positive shift indicates a shift away from the masker lateral angle. Error bars show standard error of the mean, and asterisks indicate shifts that were significantly different from zero ($p<0.05$).

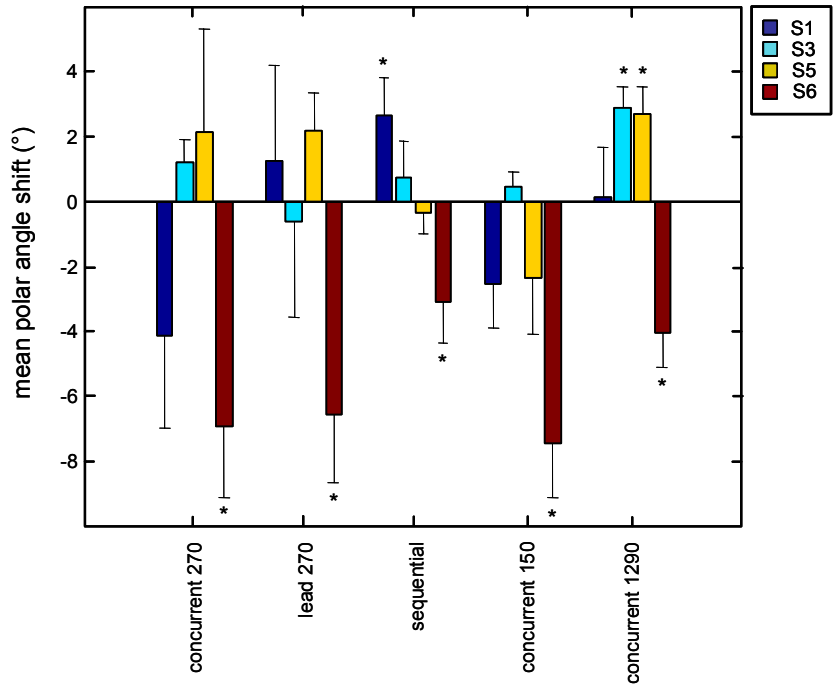



**Figure 4.9** Mean polar angle shifts (across all locations and maskers) in Experiment 1. The five groups of bars show the five stimulus conditions and each bar in a group represents a different individual. A positive shift indicates a shift away from the masker polar angle. Error bars show standard error of the mean, and asterisks indicate shifts that were significantly different from zero ($p<0.05$).

the spherical plots, and explains why little effect was seen in the previous figures where data was pooled across subjects. It appears that S6 was most consistently affected by the masker, which caused a significant negative shift in all conditions. This may indicate a bias in polar angle localisation towards the masker (pulling effect) or it may represent a bias towards the audiovisual horizon (collapse in elevation perception) as all maskers were located on this plane. The other subjects showed less consistent effects, although a positive shift was evident in S3 and S5 in the long duration condition. Note also that the sequential masker did cause a significant polar effect in two subjects (a push in S1 and a pull in S4). Overall, there were fewer *significant* shifts when compared to the lateral angle analysis.

### 4.5.3 Discussion

**Overall performance**

An important finding of this experiment was that a broadband masker did not have a striking impact on the localisation of a broadband target, despite substantial overlap in both frequency and time. This is likely a result of the good target-to-masker energy ratio employed; the target was clearly audible in all trials. A similar robustness of localisation has been shown previously with a favourable target level (e.g. Lorenzi *et al.*, 1999), although not for stimuli with simultaneous onsets and offsets. This finding demonstrates the remarkable capacity of the auditory system for extracting spatial cues from a mixed acoustic signal.

**Perception of lateral angle**

Another finding of this experiment was that a simultaneous noise masker caused a bias in the lateral angle localisation of the target. This bias was small (around 3° on average) but consistent across subjects, and was directed away from the lateral angle of the masker.

This kind of effect has been seen in a number of studies as discussed above. However, the hypothesis proposed in previous studies, that such effects are primarily a result of a masker that leads in time, is not supported by this work. In fact, we found no effect of bringing the masker on early, which is interesting because the focus of several previous studies has been on the importance of the masker lead time in localisation bias (e.g. Canévet and Meunier, 1996). It would seem from the current

results that the bias arises primarily from the *simultaneous* portion of the stimulus. It was clear that when the masker preceded the target but the two did not overlap temporally ('sequential'), there was little or no lateral bias induced by the masker. In fact localisation was affected very little by the presence of the masker in this condition.

Interestingly, previous studies using fully simultaneous stimuli have generally found that the stimuli fuse or attract each other (Butler and Naunton, 1964; Thurlow *et al.*, 1965; Heller and Trahiotis, 1996) and this is likely a grouping phenomenon. However, in the present study, the lateral angle repulsion was strong in the concurrent conditions, suggesting that grouping did not occur. Perhaps the fact that our stimuli were very distinct in terms of their temporal envelopes meant that they were streamed more effectively than those stimuli used in previous studies.

Varying the duration of a concurrent stimulus presentation did not appear to strongly alter the influence of the masker in the binaural dimension. Consistent shifts in lateral angle away from that of the masker were observed for concurrent pairs of duration 150, 270 and 1290 ms. The fact that the repulsion effect persists for the longer duration suggests that it is not simply an onset effect. It may be envisaged, for instance, that the interference or distraction due to the masker may cause an initial bias or 'overshoot' that the system might adjust to over time and arrive at an accurate estimate. This does not appear to happen in the span of 1290 ms, as the lateral push is still evident in the data.

**Perception of polar angle**

The effect of a masker on polar angle localisation was found to be non-systematic and strongly individualised. In the concurrent conditions, there did seem to be a tendency for polar angle to be 'pulled' towards 0°, especially in S6. However in the long duration condition a significant push was observed in two subjects.

This study was one of only a few that have examined localisation bias in the vertical (or polar angle) dimension, and thus comparisons with the literature are difficult. Getzmann (2002; 2003b) used an experimental set-up where the masker lead by 1 second. This author reported a consistent bias away from the masker, with a magnitude of around 3-5°. In the present 'lead 270' condition, which is perhaps the most comparable, two subjects did show pushing effects, but these were not significant. In addition, the other two subjects showed a bias in the opposite direction

on average (away from the masker). Getzmann (2003b) also observed a pushing effect when target and masker were presented sequentially. This bias was about one third as large as the bias in the concurrent condition. This effect is similar to the present findings in the lateral dimension, where sequentially presented target/masker pairs produced a much reduced bias in target localisation. However the polar angle results in the present study were too variable to draw such conclusions.

**Questions raised by Experiment 1**

A major finding from Experiment 1 was that a completely simultaneous masker had a reasonably minor effect on the localisation of the target sound source. This result was somewhat surprising as previous studies have demonstrated significant effects of a masker on target localisation, including confusions between the two, and an integration of spatial information to give estimates biased towards each other. These larger effects are not unexpected, as interference between simultaneous auditory objects is inevitable based on the non-spatiotopic design of the auditory periphery. One possible explanation of the lack of major localisation deficits observed in Experiment 1 is that the target was strongly dominant over the masker and the influence of the masker on spatial perception was essentially swamped. Although the two stimuli were RMS-matched overall, the noise-train was indeed more intense at the *peak* of its modulations than the steady masker. To investigate this notion, the 'concurrent 270' condition was repeated in Experiment 2 with the target salience reduced in two ways (see section 4.6.1).

Another finding from Experiment 1 was that the masker tended to repel the target resulting in localisation biases. As this effect was highly consistent in the horizontal dimension, it was of interest to determine whether it might be attributable to effects on one of the binaural cues (ITD or ILD). A third condition was included in experiment 2 as a preliminary means of investigating this question. Stimuli were high-pass filtered in order to remove dominant low-frequency ITD information, thus reducing the binaural spatial cue set. An attempt was also made to repeat the experiment using low-pass filtered stimuli, in order to remove the strong ILD cues available in the high-frequency region. Such stimuli were used with success in the two-point discrimination experiment of Chapter 3. However, these low-frequency sounds are poorly localised (recall section 1.4.1) and most of the subjects in this study did not have a good externalised percept of these sounds. Thus it was impossible to

obtain reliable enough estimates of perceived location to enable the kinds of analysis required.

## 4.6 Experiment 2: Influence of stimulus characteristics

### 4.6.1 Stimuli

Experiment 2 consisted of three stimulus conditions (Figure 4.10). In all conditions, the stimuli were simultaneous with a total duration of 270 ms (similar to condition A of Experiment 1), but the characteristics of the stimuli were modified. The first two conditions addressed the issue of whether the noise-train target in Experiment 1 might have been dominant over the continuous noise masker. In condition A ('inverse'), the target and masker were simply interchanged, so that the continuous noise was elected as the target to be localised. In condition B ('reduced TMR'), the target and the masker were not RMS-matched as in the previous conditions. Instead, the target level remained the same as in previous conditions, but the masker level was increased such that it approximately matched the *peak amplitude* of the target. In comparison to the previous arrangement, this meant increasing the masker level by approximately 7 dB, resulting in a new target-to-masker (TMR) ratio of approximately –7 dB. The percept then was of a far less dominant, but still salient, target. In condition C ('high-pass'), stimuli were identical to the 'concurrent 270' condition of Experiment 1 but were high-pass filtered at 2 kHz. Filtering was performed in the time domain using brick-wall FIR filters (60 dB down in the stop band).

### 4.6.2 Results

**Overall performance**

The SCCs for control and test conditions can be found in Table 4.2. In the control (no masker) trials, SCCs ranged from 0.59 to 0.95. Notice that S1 showed a drop in accuracy when the noise-train was high-pass filtered ('high-pass') and clearly found the continuous noise harder to localise than the noise-train ('inverse'). This latter effect has been observed in listeners of other studies (e.g. Getzmann, 2003a).

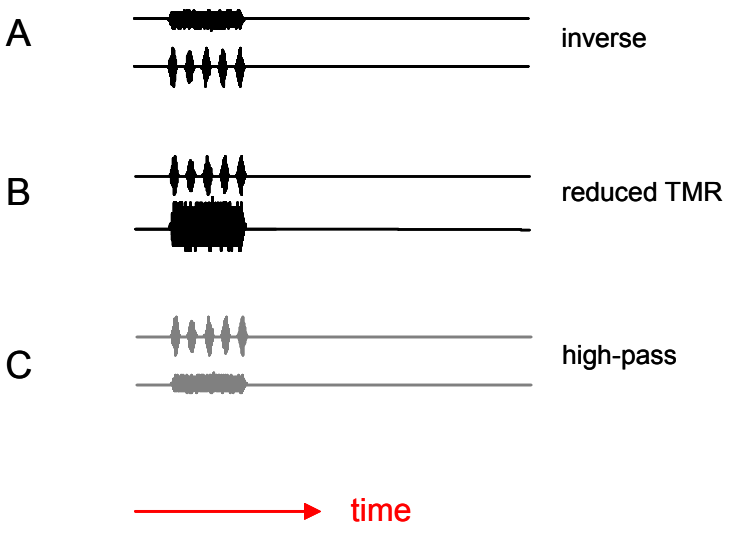



**Figure 4.10** The three stimulus conditions of Experiment 2. All are based on the 'concurrent 270' condition of Experiment 1. Condition A('inverse'): the target and masker were interchanged. Condition B ('reduced TMR'): the masker level was increased such that the peak amplitudes of target and masker were matched. Condition C ('high-pass'): both target and masker were high-pass filtered at 2 kHz (indicated by grey colour).

When presented in conjunction with the masker, target localisation was relatively well maintained. SCCs ranged from 0.62 to 0.93, and were similar in general to the corresponding control conditions. Interestingly, S1 appeared to *improve* in localisation of the continuous noise ('inverse') when the noise-train masker was present.

**Table 4.2** Spherical correlation coefficients (SCCs) for each of the four subjects (S1, S3, S5, S6) in control (no masker) trials and test (with masker) trials of Experiment 2. Each of the three rows contains values for the three stimulus conditions. Each SCC was calculated from 175 control trials or 285 test trials. See section 2.4.3 for details of this statistic.

| | | S1 | S3 | S5 | S6 |
|---|---|---|---|---|---|
| inverse | cont | 0.59 | 0.92 | 0.94 | 0.91 |
| | test | 0.70 | 0.93 | 0.93 | 0.81 |
| reduced TMR | cont | 0.93 | 0.94 | 0.93 | 0.90 |
| | test | 0.79 | 0.90 | 0.85 | 0.62 |
| high-pass | cont | 0.76 | 0.95 | 0.92 | 0.85 |
| | test | 0.70 | 0.93 | 0.90 | 0.71 |

### Qualitative analysis of responses

Figures 4.11-4.13 contain spherical plots of the results from the three conditions. The details of these figures are the same as for Figures 4.3-4.7. The four rows depict results from the four subjects, and the three spherical plots in a row show data for the three masker locations (left: 30°,0°; middle: 0°,0°; right -30°,0°). Again the masker location is depicted by a large red dot and the location directly in front of the listener is indicated by the grey circle or 'nose'. Centroids of localisation estimates in the presence of the masker (blue dots) are connected by blue lines to their corresponding control centroids (red dots).

In the 'inverse' condition (Figure 4.11), the masker had a noticeable effect on elevation perception only in S1 and S6. This emerged as an upward bias in elevation for S1, but a general collapse towards 0° elevation for S6 (which may indicate a 'pull' towards the masker polar angle). Lateral estimates were affected in all subjects, and in many cases this was bias away from the lateral angle of the masker. However, this effect was less consistent than seen in previous conditions (Experiment 1).

In the 'reduced TMR' condition (Figure 4.12) the masker produced a greater disruption in localisation than in previous conditions. It appears that localisation estimates in the vicinity of the masker are clustered at or near the location of the masker, causing disruptions in both lateral and polar angle. This suggests that the target may have been fused with the masker in these instances. For locations more distant from the masker, target localisation is less affected. However, a lateral angle bias away from the masker is apparent, and it appears to be larger in magnitude than seen in previous conditions.

A similar clustering of responses at the location of the masker was observed in S6 for the 'high-pass' condition and to some extent in S1 (Figure 4.13). Away from the masker location, lateral angle 'pushing' effects were again apparent. S3 and S5 also showed evidence of lateral angle pushing, as well as some relatively minor disturbances to elevation responses.

### Quantification of localisation bias

The difference between masker and no-masker centroids was calculated and expressed as lateral and polar angle shifts in the same way as in Experiment 1. Data from all masker locations was pooled, but individual subjects were treated separately.

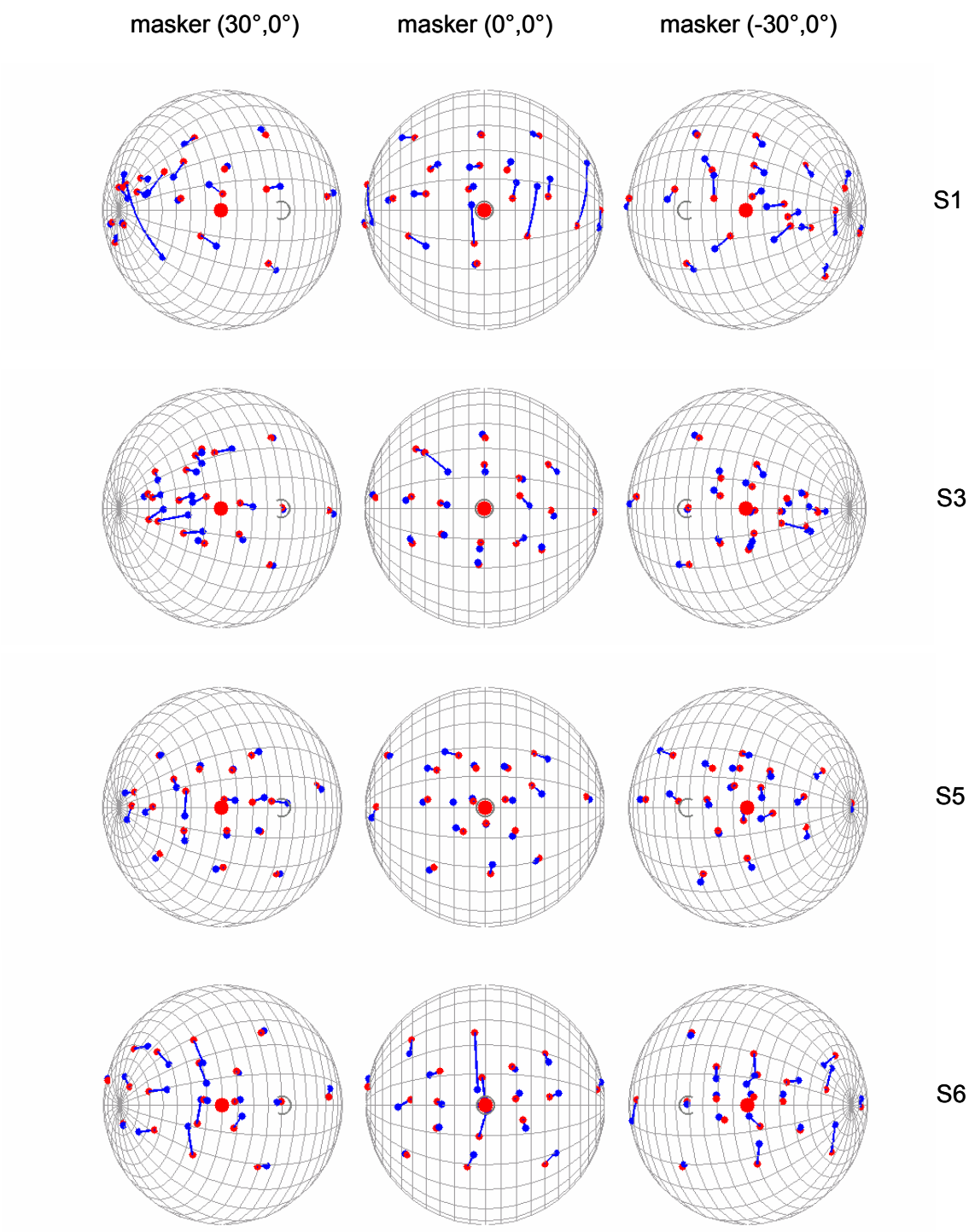



**Figure 4.11** Spherical plots for the 'inverse' condition of Experiment 2. All other details as for Figure 4.3.

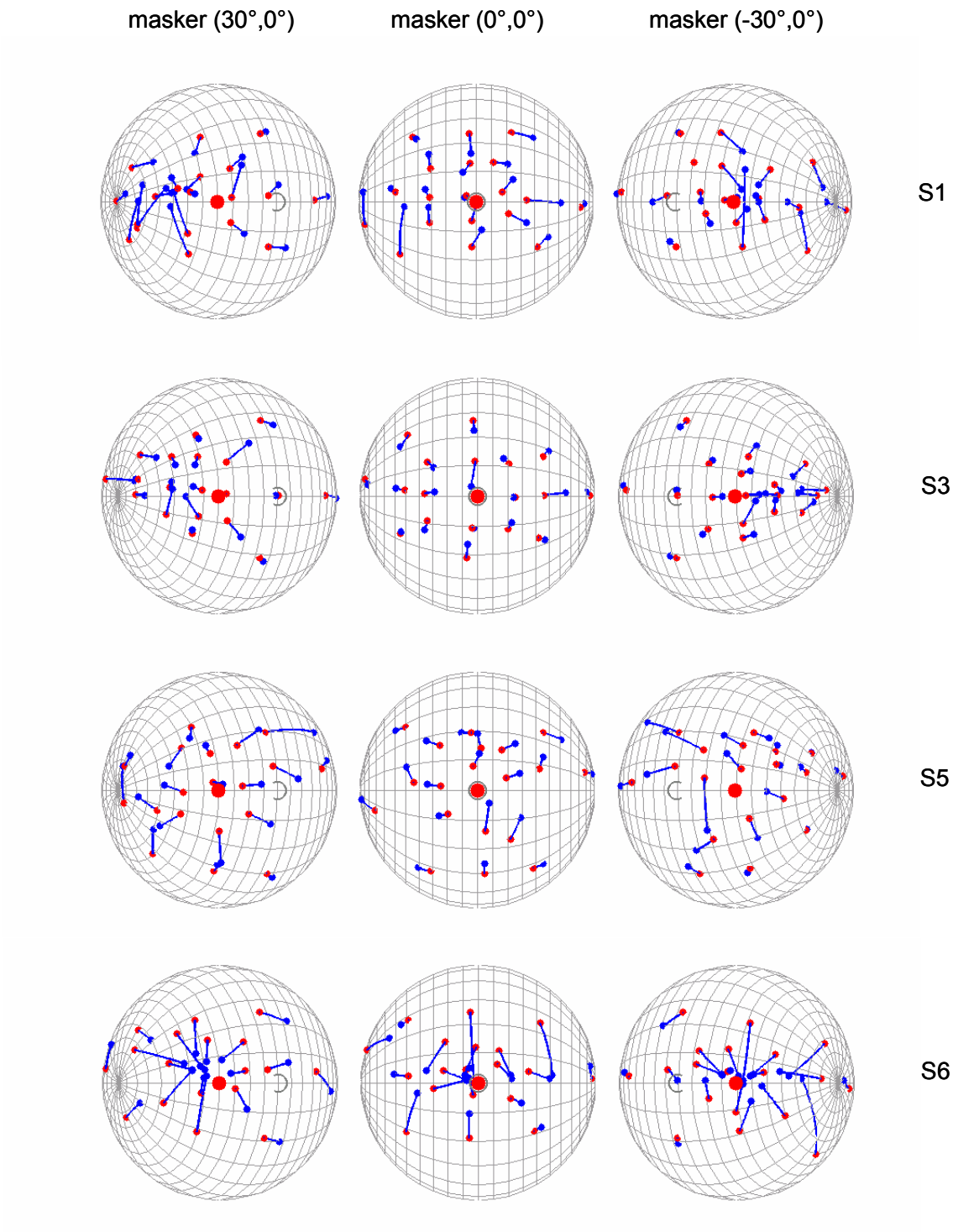



**Figure 4.12** Spherical plots for the 'reduced TMR' condition of Experiment 2. All other details as for Figure 4.3.

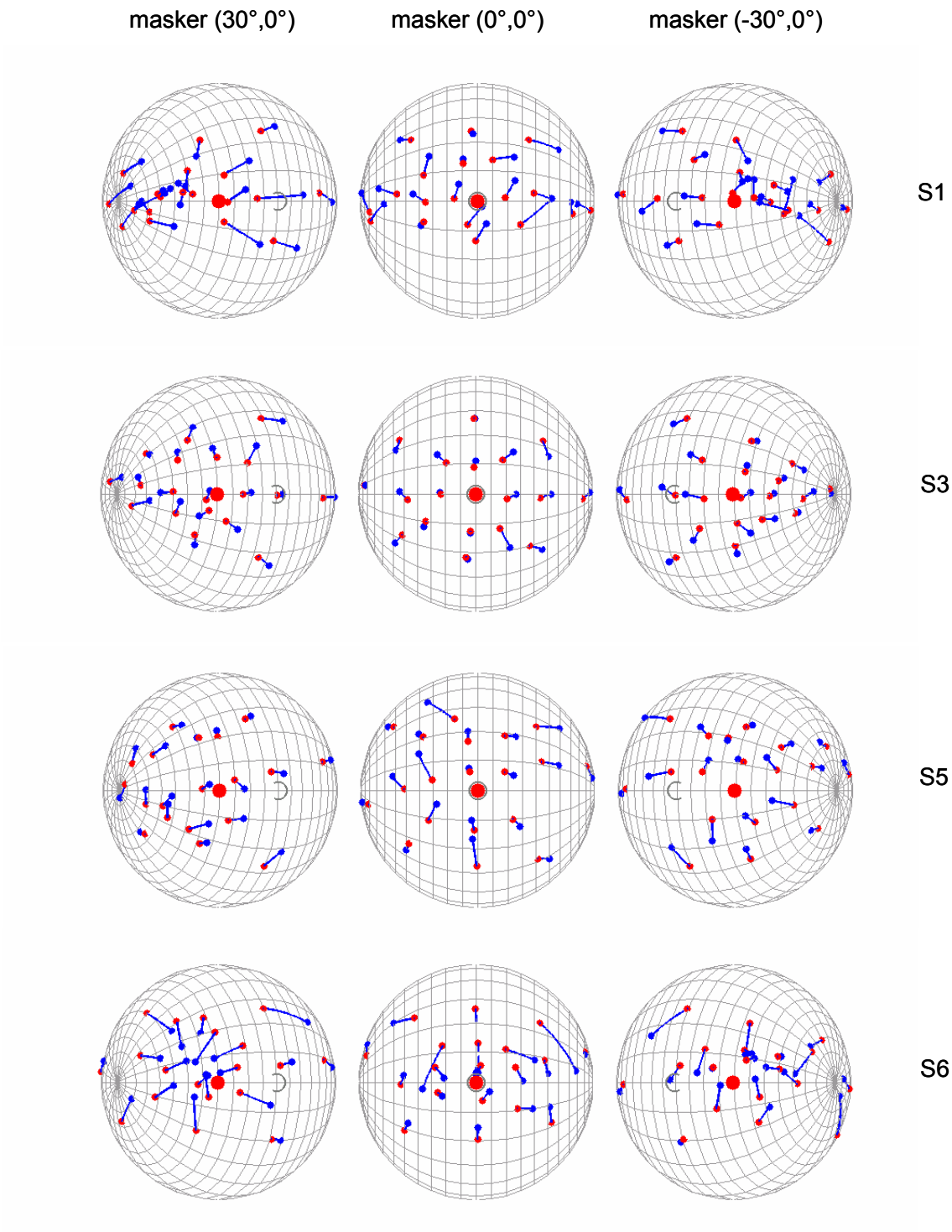



**Figure 4.13** Spherical plots for the 'high-pass' condition of Experiment 2. All other details as for Figure 4.3.

Again, a positive shift indicates a bias away from the masker and a negative shift indicates a bias towards the masker.

Figure 4.14 illustrates mean lateral angle shifts. The three clusters of bars represent the three conditions, and the individual bars represent the individual subjects. Note that all mean lateral angle shifts (except one) are positive, indicating a bias away from the masker lateral angle. The error bars show the SEM and the asterisks indicate shifts that were significantly different from zero (two-tailed t test, $p<0.05$).

For the inverse condition, it can be seen that biases were small, and only significant in one subject (S3). This confirms the observation made above, that shifts were smaller than in previous conditions. In fact shifts were approximately 1° on average in this condition, where the continuous noise was the target. In the equivalent condition in Experiment 1, in which the target and masker were exchanged ('concurrent 270'), shifts were on the order of 3°. The lateral angle shifts for the 'reduced TMR' condition vary between the four subjects. Three subjects show a positive shift, and in S5 it is particularly large (5.6°). In this subject and S1 the shift is significant. S6 shows a non-significant pull on average, which is in contrast to 'concurrent 270' condition in which she showed a push of nearly 2.5°. In the 'high-pass' condition, all subjects showed a significant positive shift reminiscent of that seen in Experiment 1 (approximately 3° on average).

The results from the 'concurrent 270' condition of Experiment 1 were included in the statistical analysis of trends. An ANOVA on the group data revealed a significant effect of condition on the lateral angle shift [$F(3,908) = 5.17$, $p = 0.002$]. Post-hoc analysis (Tukey HSD, $p=0.05$) showed that the 'inverse' condition was significantly different from the 'concurrent 270' condition, suggesting that using the continuous noise as the target reduced the lateral bias. Overall, the 'reduced TMR' results were not different from the 'concurrent 270' condition, in which the TMR was more advantageous. Furthermore, the 'high-pass' results were not significantly different from the broadband condition ('concurrent 270').

Figure 4.15 illustrates the mean polar angle shifts. Again, each cluster of bars represents one of the three conditions, the individual bars represent the individual subjects, the error bars show SEMs, and the asterisks indicate significant shifts. Interestingly, all subjects demonstrate a negative polar angle shift in these conditions, indicating an overall pulling towards the 0° elevation plane on which the maskers

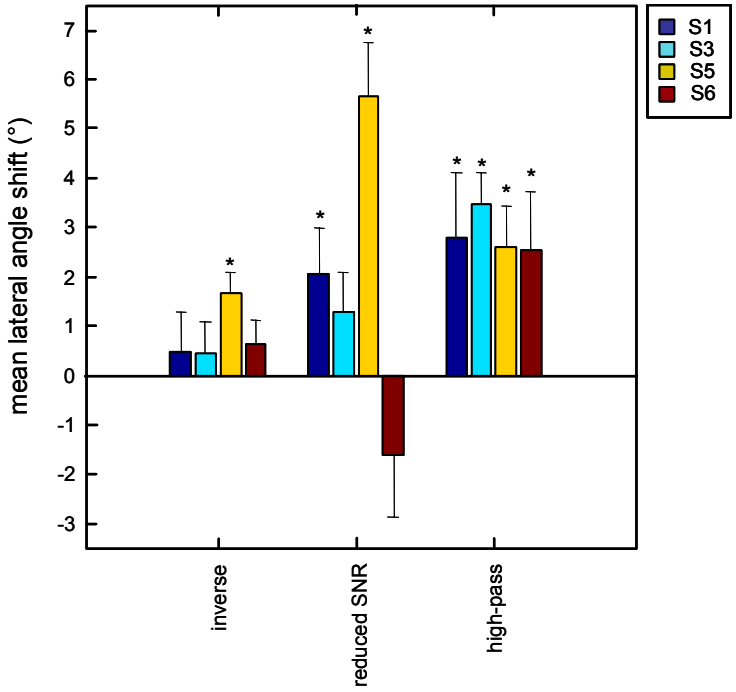



**Figure 4.14** Mean lateral angle shifts (across all locations and maskers) in Experiment 2. The three groups of bars show the three stimulus conditions and each bar in a group represents a different individual. A positive shift indicates a shift away from the masker lateral angle. Error bars show standard error of the mean, and asterisks indicate shifts that were significantly different from zero ($p<0.05$).

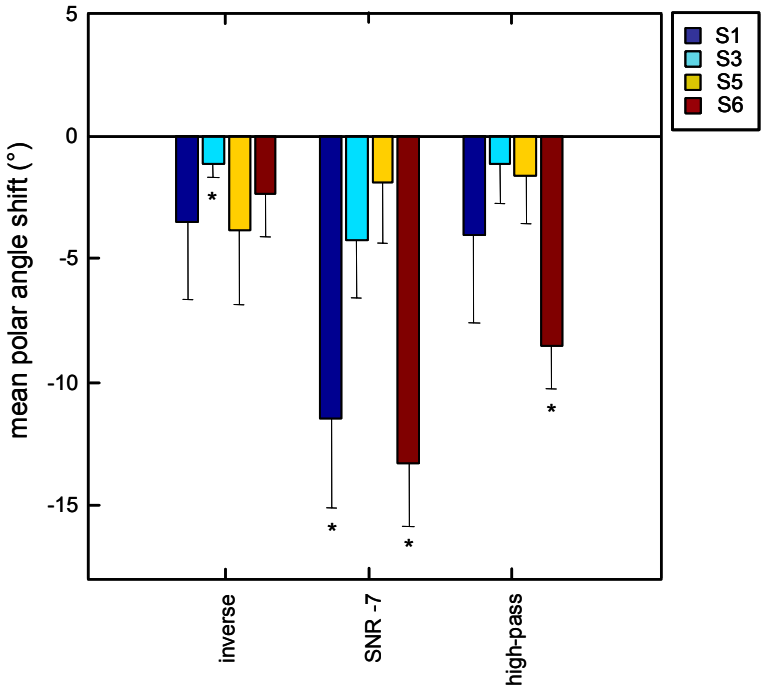



**Figure 4.15** Mean polar angle shifts (across all locations and maskers) in Experiment 2. The three groups of bars show the three stimulus conditions and each bar in a group represents a different individual. A positive shift indicates a shift away from the masker polar angle. Error bars show standard error of the mean, and asterisks indicate shifts that were significantly different from zero ($p<0.05$).

were located. Recall that in Experiment 1, S6 was the only subject who consistently showed this sort of negative shift. For the 'inverse' condition, the shift was only significant for S3. For the 'reduced TMR' condition, the pull is extremely strong (and significant) in S1 and S6. This is consistent with the spherical plots (Figures 4.11-4.13), which showed these two subjects to most strongly cluster targets to the masker location. In the 'high-pass' condition, S6 showed a significant pull in polar angle, while mean shifts in the other three subjects were non-significant. Note that this pattern is very similar to that seen using broadband stimuli (Experiment 1, 'concurrent 270').

### 4.6.3 Discussion

**Effect of reducing the dominance of the target**

The 'inverse' condition was employed to examine whether the noise-train may have dominated the continuous noise in Experiment 1, giving results that may not extend to other stimulus combinations. To test this, the target and masker were interchanged, and subjects localised the continuous noise. Overall, the *pattern* of results was not markedly different from the equivalent condition in Experiment 1 where subjects localised the noise-train ('concurrent 270'); in both conditions, localisation was reasonably good in the presence of the masker, but a consistent lateral angle bias was observed that was directed away from the masker. In the 'inverse' condition, however, this bias was smaller (about one third of the magnitude) than in the 'concurrent 270' condition and only significant in one subject.

Recall that in Experiment 1 it was concluded that the simultaneous portion of these target/masker stimuli is responsible for most of the lateral angle bias. In the 'inverse' condition, the target is actually on for very short periods *without* the masker (in the gaps of the noise-train). If we imagine that localisation of a prolonged sound involves taking successive estimates or samples throughout its duration, then these gaps might give opportunities for localisation of the target alone, without interference from the masker. Then when successive estimates are integrated into a single percept, the inclusion of 'pure' location estimates could certainly reduce the bias induced by the masker. A similar argument was made in a recent study by Getzmann (2003a). He examined what he terms 'contrast effects' and found that in the horizontal plane, a noise target is repulsed less by a noise-train masker than a continuous noise masker.

He suggested that stable estimates of the target location could be obtained in the (50 ms) gaps of the noise-train masker. Indeed it is known that these gap durations (50 ms in Getzmann's study and 30 ms in the present study) are long enough for accurate binaural estimates (Hofman and Van Opstal, 1998).

When the original stimulus arrangement was used, but the TMR was reduced from 0 dB to –7 dB, an interesting effect was observed. In the vicinity of the masker, subjects tended to cluster target responses at (or near) the location of the masker, producing a pull in both lateral and polar angle. It seems that when the target and masker occupied similar locations, the target was so overpowered by the masker that it fused with it. However, when the target was more distant from the masker, this effect was not apparent and responses looked similar to that seen in previous conditions. This result suggests that there is a boundary for segregation and fusion, and this idea was addressed in the previous Chapter. Clearly for objects close together there is a point at which they will fuse perceptually. However the present results suggest that this boundary is driven by the relative levels of the two objects (the TMR in this case) and the dominant object will tend to drive localisation when fusion does occur.

**Effect of high-pass filtering**

The experiment was conducted using high-pass filtered stimuli in order to begin to probe the spatial cues involved in the observed localisation bias. By high-pass filtering, the low-frequency ITD cue was removed. The results showed that localisation in the presence of a concurrent masker was subject to the same effects seen with broadband stimuli. In particular, the systematic lateral bias was still apparent. Thus, it is unlikely that the bias is due to interference in the extraction of ITD, as removal of the primary ITD cue did not reduce the effect. However, it is known that there is high-frequency ITD information in the envelope, and as described in the previous chapter, it can provide useful information in a concurrent source situation. Furthermore, the envelope of the noise-train target in the present study is particularly well-defined due to its deep modulations. Thus, the role of ITD processing cannot be excluded, and is discussed further below (section 4.7.3). Alternatively, the high-pass results are also consistent with a mechanism involving ILD processing, an idea also explored below.

# 4.7 General discussion

## 4.7.1 Comparison with the literature

Overall, these experiments were informative in terms of reconciling the seemingly conflicting results reported in the literature on this topic. Firstly, the fact that subjects in the present study localised reasonably well in all conditions is consistent with those previous studies that reported that a masker had little or no effect on target localisation (e.g. Lorenzi *et al.*, 1999). Secondly, in one case, clear pulling effects of the masker were seen, an effect that has been observed in other studies (Butler and Naunton, 1964; Thurlow *et al.*, 1965; Heller and Trahiotis, 1996). Finally, the lateral angle pushing effects seen by other authors (e.g. Braasch and Hartung, 2002; Canévet and Meunier, 1996) were replicated, and shown to be due to simultaneous portions of the presented stimuli. It is worth mentioning that this effect is truly a *perceptual* bias and not simply a *response* bias, because verbal position estimates have been shown to give similar results (see Getzmann, 2003b; Bridgemann *et al.*, 1997).

## 4.7.2 Relation to models for localisation bias

Several authors have proposed models to explain the repulsion of perceived target location by a masker. Many of these models are based on the idea of auditory 'spatial channels', where stimuli from different spatial locations can influence each other via a neural array. It is generally assumed that in such an array, neighbouring locations in space are represented by neighbouring neurons (spatiotopic map) or neurons having strong connections with each other. Carlile and colleagues (Carlile *et al.*, 2001) proposed a model based on such an array to explain localisation bias following exposure to a prolonged stimulus at a particular spatial location. In this model, adaptation to a prolonged stimulus results in a down-regulation of spatial channels tuned to that location, giving an imbalance in the excitability of the array. Thus subsequent stimuli presented at locations close to the adapted region are localised *away* from the region of reduced excitability (Figure 4.16a). This kind of model is useful for explaining localisation bias caused by prior exposure to a masker. It may begin to explain the small pushing effect seen in two subjects in the sequential

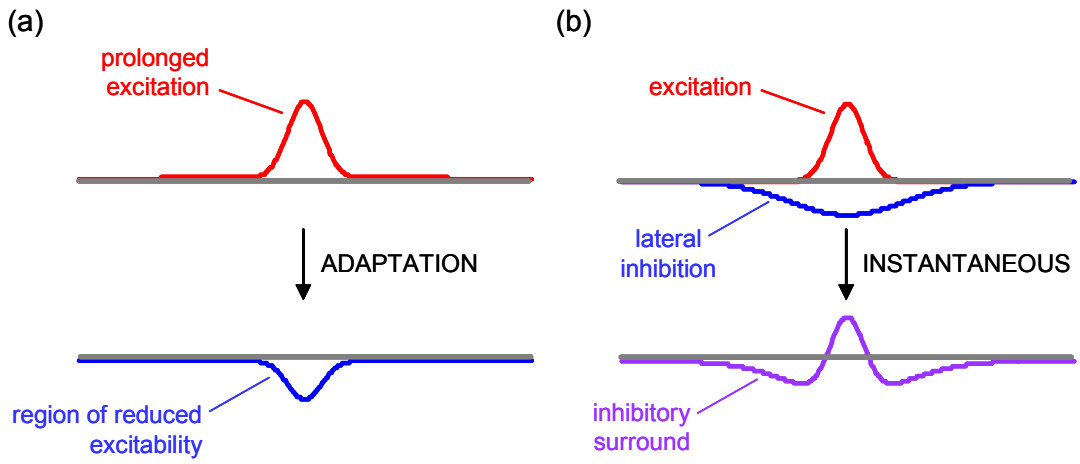



**Figure 4.16** (a) Model for adaptation to a spatial location. Prolonged excitation of the neural array at a particular location (red) causes a down-regulation in the channels tuned to that location (blue). Subsequent stimuli presented at locations close to the adapted region may be localised *away* from this region of reduced excitability. (b) Model for instantaneous lateral inhibition effects. In this common model, activation of a particular spatial channel (red) causes a fast down-regulation of neighbouring channels via lateral inhibition (blue). This produces a region of reduced excitability (purple), away from which stimuli in close proximity may be repelled.

condition of the present study, and similar effects reported by others using non-simultaneous stimuli (Getzmann, 2003b; Kashino and Nishida, 1998). However, as reported in the present study and that of Getzmann (2003b), localisation bias can be much larger for stimuli that overlap in time, requiring an explanation other than adaptation.

A few authors who have observed pushing effects in the horizontal plane with concurrent stimuli have proposed that a sensory contrast mechanism may be able to explain the effects. This model is similar to that described above for adaptation effects, but it proposes that spatial channel interactions may be instantaneous. Suzuki *et al.* (1993) describe a form of 'spatial masking', where activation of a particular spatial channel causes a down-regulation of neighbouring channels (Figure 4.16b). Indeed in vision this kind of lateral inhibition mechanism is known to enhance spatial contrast to aid in edge detection, and in fact lateral inhibition is a general sensory principle at work even in olfaction (Aungst *et al.*, 2003).

However, a model based on interactions within spatial channels would predict that neighbouring locations would have the largest impact on each other, and that interference would decrease with increasing distance between the target and masker.

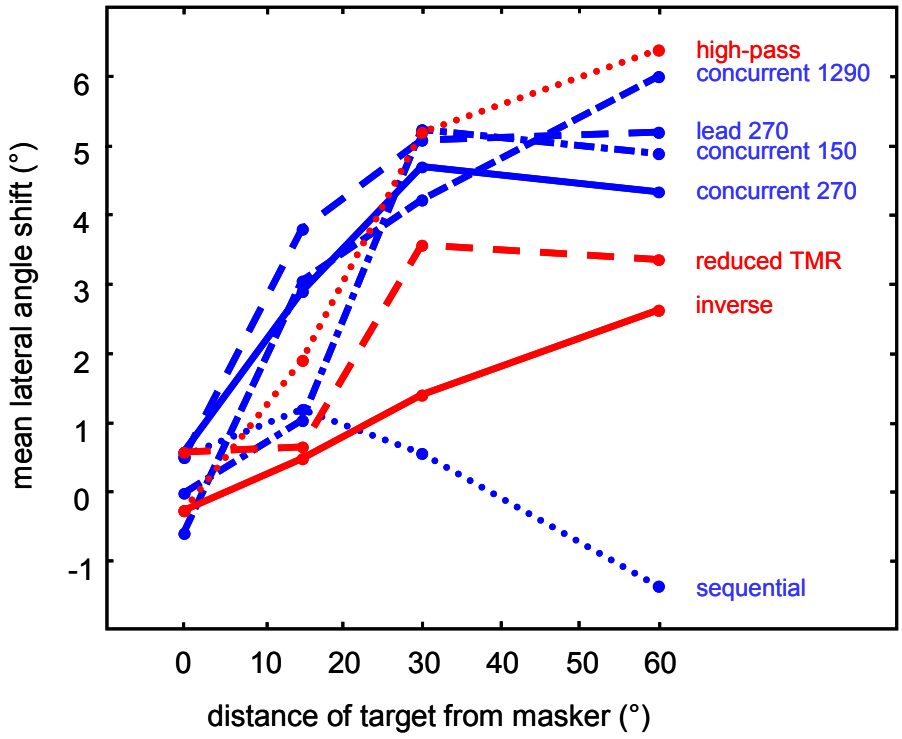



**Figure 4.17** Effect of distance from the masker on lateral angle localisation bias. Responses have been pooled across subjects and locations and the mean lateral angle shift is plotted against the distance (in lateral angle) of the target from the masker. The blue lines show the five conditions of Experiment 1 as labelled, and the red lines show the three conditions of Experiment 2 as labelled.

In Figure 4.17, the data from all subjects has been pooled and mean lateral angle shifts have been plotted as a function of absolute distance of the target from the masker. The five conditions from Experiment 1 are shown in blue, and the three conditions from Experiment 2 are shown in red. It can be seen that for all concurrent conditions, the closest target locations were subject to the bias, but that the more distant locations were subject to a much greater bias. Thus a model based on local channel interactions is not supported by this data.

The localisation disturbances observed in the present study in the vertical dimension are also not consistent with a simple spatial channel model. Distortions to elevation perception were highly non-systematic and varied greatly in magnitude. This is in contrast to a recent study by Getzmann (2003b), who reported systematic elevation distortions for concurrent stimuli. However Getzmann's bias was asymmetric for simultaneous sounds. Together, these results demand an explanation other than spatial channels or spatial contrast, as these mechanisms would predict a

systematic and symmetric effect. An acoustic explanation may be more apt to explain these idiosyncratic effects (see below).

It was very interesting to note that the results in the 'sequential' condition were reasonably compatible with a spatial channel model. In Figure 4.17, it can be seen that for this condition alone, lateral shifts on average are largest for target locations adjacent to the masker. This is consistent with the study of Carlile and colleagues where this kind of effect was seen for sequential stimuli, although in that case the masker was greatly prolonged (Carlile *et al.*, 2001). The bias observed by Carlile and colleagues also extended to the vertical dimension, but a systematic vertical bias was not evident in the present data. This may be related to the brevity of the masker in this experiment.

### 4.7.3 Segregation and localisation cue processing

So, while channel-based interactions may have an effect in multiple-source situations, it is clear that a more dominant effect is acting when stimuli are presented simultaneously. Presumably the auditory system must effectively segregate competing signals in order to localise one (or several) of them, and if this segregation is imperfect, then it can be expected that localisation will be disrupted. In the present study, stimuli overlapped substantially in frequency and time, making segregation and independent localisation a difficult task for the auditory system. Indeed subjects remarked that the target noise-train sounded 'noisier' when presented in conjunction with the noise masker, suggesting that although two auditory objects were perceived, the segregation was not complete.

Importantly, acoustic interference of this kind could differently affect binaural and spectral cue processing. Thus, this is an appealing basis for explaining the different effects observed in the present study for horizontal and vertical localisation.

**Interference in binaural processing**

There is considerable evidence showing that ITD processing is disrupted by competing binaural signals *only* if segregation of target and masker is impaired. ITD discrimination using tonal stimuli is disrupted more by the presence of a harmonic masker than an inharmonic masker (Buell and Hafter, 1991). In addition, ITD detection in a narrow frequency band is impaired when flanked by a diotic broadband

noise, but only if the noise is gated on and off together with the target band (Trahiotis and Bernstein, 1990). Interestingly, Zurek (1985) saw that ITD sensitivity for high-frequency signals and not low-frequency signals was reduced in the presence of a broadband masker. As the lateral effects in the present study were evident for broadband and high-pass filtered stimuli, this kind of disruption to high-frequency ITD could play a role. It is also possible that low-frequency ITD effects play a role, although it was not possible in the present study to examine this idea

Several other authors have used low-frequency ITD-based mechanisms to explain localisation bias in the horizontal plane. Kashino and Nishida (1998) explained localisation bias induced by prolonged exposure to an adapting sound by introducing gain control step into a cross-correlation model. Braasch (2002) used a similar approach to explain localisation bias in the case where a masker preceded a target by only 200 ms. This kind of explanation is appealing for describing the present results. However, the model requires that there is an interval in which the masker is present on its own. This requirement was explicitly met in one condition of the current study ('lead 270'). Furthermore, in most of the conditions there were 30 ms gaps within the noise-train target in which the masker may have been sampled. However, in the 'inverse' condition the masker was never presented in isolation. Furthermore, an identical lateral bias was observed using concurrent monosyllabic speech stimuli where both sounds were continuous (see Chapter 6). So while an ITD-based model has been successful in explaining many aspects of localisation bias, it must be modified to cater for simultaneous effects, and furthermore must be extended to operate in high-frequency ranges.

Only one study was found in which ILD rather than ITD processing in competing source situations was considered. In this study (originally published in Japanese by Itoh and colleagues, cited in Suzuki *et al.*, 1993) the lateral position of auditory images based on ILD was examined in the presence of a masker. Results showed that the lateral image moves outwards if a centred (zero ILD) stimulus is presented simultaneously. Thus it is possible that the effects seen in the present study are related to a disruption in ILD processing. It may be that the auditory system extracts the target signal from the masker signal at each ear independently, and it is conceivable that this segregation is incomplete for stimuli of similar spectral content. Furthermore, this segregation may be unequal in the two ears, with less signal recovered from the 'poorer' ear (where the masker is more intense relative to the

target). Such an imbalance would mean that an ILD estimate based on the recovered signals would be biased in a consistent way away from the masker ILD.

The results presented in this chapter are consistent with an ILD-based explanation, as high-pass filtered stimuli were as strongly affected as broadband stimuli. Furthermore, inspection of the spherical plots (Figures 4.3-4.7 and 4.11-4.13) reveals that the lateral bias was notably strong for target/masker pairs located at ±30°. This arrangement places the stimuli approximately at the acoustic axes (positions of maximum gain due to the directionality of the pinnae), making ILD disturbances most likely.

**Interference in spectral analysis**

It is more complicated to consider interference that may occur in the extraction of monaural spectral cues in a concurrent source situation. The fact that listeners in the present experiment were able to make reasonable polar angle judgements in the presence of a masker suggests that the auditory system has some ability to recover the spectral cues corresponding to a target, despite the fact that they are confounded with the spectral cues of the masking noise. It is not at all clear how the auditory system performs this non-trivial task. Simple models based on HRTF template matching have not been very successful (Langendijk *et al.*, 2001) and a more sophisticated model is required.

The fact that polar angle effects in this study were highly dependent on the locations of the target and the masker suggests that there is interference in the extraction of spectral cues. Different combinations of target and masker location will produce a different composite spectrum at the ear, and imperfect segregation would produce two spectra varying from the originals. As the spectral cues vary in a complex way across locations, spectral changes can result in idiosyncratic changes in perceived location. Furthermore, the resultant spectra will be processed by the system in a way that depends on knowledge about the two stimuli and learned associations related to the acoustic properties of the individual's head and pinnae. This may explain the highly individualised nature of polar angle biases observed in the present study.

### 4.7.4 A note on attention

It must be pointed out that the role of attention in this kind of task has not been discussed but is certainly of importance. Discussions with participants after the experiments revealed that they had some sense of the location of the masker but that it was often far from the region in which the masker was actually located. While all stimuli were positioned in the frontal hemisphere, and targets were always correctly localised here, the masker was often perceived to be located behind or above. According to Wightman and Kistler (1992), there is relatively little variation in the head-related transfer functions in these regions. According to their observations, stimuli containing sub-optimal localisation cues are often localised to such regions of naturally high uncertainty. Perhaps the loss of spatial accuracy for non-attended stimuli is due to a 'perceptual attenuation', as described by Botte and colleagues (Botte *et al.*, 1997). These authors observed that non-attended auditory streams need to be presented at a higher intensity level than attended streams to enable the detection of temporal irregularities. If such a perceptual attenuation of the non-attended stimulus occurred in the present experiment, it is reasonable to expect that spatial cues would not be extracted efficiently. Thus it would seem that when there are competing auditory objects, a particular object can be localised to a reasonable degree of accuracy, but *only* if it is attended.

## 4.8 Conclusions

This experiment investigated the effect of a simultaneous broadband masker on the localisation of a broadband target stimulus. It was found that, overall, localisation was quite robust. However, small disruptions in target localisation were observed for cases in which the target and masker were played simultaneously. These effects included a bias in lateral angle perception that was systematically directed away from the lateral angle of the masker. Disruptions to polar angle localisation were far less systematic and highly individualised.

Previous studies have explained such results in terms of a distortion of perceptual space that is induced by interactions within a neural map of space. Such interactions may play a role in multiple-source situations and appear to be relevant for

non-simultaneous objects. However, for completely simultaneous sound sources of similar spectral content, it is likely that there is disruption to the extraction of the relevant acoustic localisation cues. Lateral angle effects may be explained by interference in ITD and/or ILD processing, and polar angle effects may be explained by impaired extraction of the appropriate spectral cues.

# Chapter 5: Speech localisation

## 5.1 Introduction

Human speech is a dynamic acoustic stimulus, varying in both frequency and intensity over time in a complex manner. It is best described in the form of a spectrogram, an example of which is provided in Figure 5.1 (Moore, 1997). In this display, energy levels in different frequency bands (ordinate) over time (abscissa) are depicted. The darker areas represent high-energy regions, and in this example a range of different speech patterns can be seen. Notice particularly (a) the vertical striations representing the periodicity of the vocal fold vibrations, (b) the horizontal bands representing the formants, which are important in defining vowel sounds, (c) the predominance of energy at frequencies below about 4 kHz, and (d) the region of broadband energy representing the "s" sound.

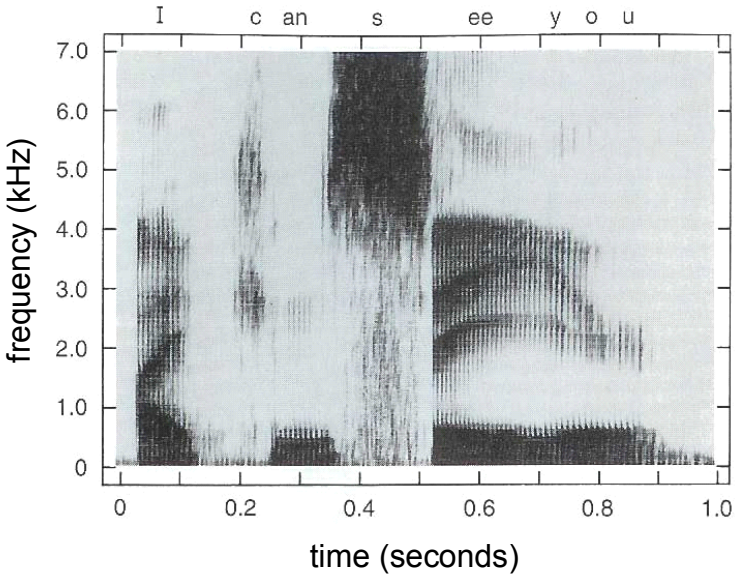



**Figure 5.1** A spectrogram of natural speech (the phrase "I can see you"). *From Moore, 1997, p. 278.*

Speech recognition depends on the identification of acoustic patterns, but is also influenced strongly by contextual cues, grammatical rules, the acoustic environment, and a listener's familiarity with the language. Furthermore, in natural

communication situations, the interpretation of speech is influenced by the style of delivery (volume, speed, emphasis) and visual cues (lip-movements, facial expressions, gestures). For these reasons, speech is ultimately processed in unique and specialised higher brain centres. However it is initially encoded by the auditory system in ways common to all acoustic stimuli. Auditory nerve responses to speech sounds, for example, can generally be explained on the basis of what is known of simple sounds (see Chapter 9 in Pickles, 1988)

It may also be assumed that the spatial processing of speech signals is no different from the spatial processing of any other acoustic stimulus (see Chapter 1). However, the dynamic nature of speech must provide the spatial auditory system with particular challenges. In particular, it is likely that spectral fluctuations inherent in the source spectrum would hinder the extraction of the spectral cues introduced by the head and pinnae.

There has been little attention given in the literature to how accurately human listeners localise speech sounds. However, speech is widely used as a stimulus in auditory research, especially in the important areas of hearing loss and communication in noisy environments. Importantly, it has become apparent that such studies have frequently used low-pass filtered (8 kHz or below) speech stimuli. The explanation for this is that most of the important sound components for speech recognition, such as formants, occur well below 8 kHz. Indeed standard clinical auditory examination involves pure tone threshold testing at frequencies from 125 Hz to 8 KHz.

However, naturally produced speech does contain significant amounts of energy above 8 kHz. In Figure 5.1 it can be seen that the broadband region representing the “s” sound is extremely strong at the 7 kHz cut-off point. In fact, broadband segments such as this can extend up to around 20 kHz (an example is shown in Figure 5.2b). High-frequency content, and certainly frequencies above 8 kHz, are important for accurate auditory localisation (Carlile and Delaney, 1999; King and Oldfield, 1997; van Schaik *et al.*, 1999; Bronkhorst, 1995). As the impact of these high-frequency cues on speech localisation in particular is unclear, one aim of the current experiment was to compare localisation of broadband and low-pass filtered speech stimuli.

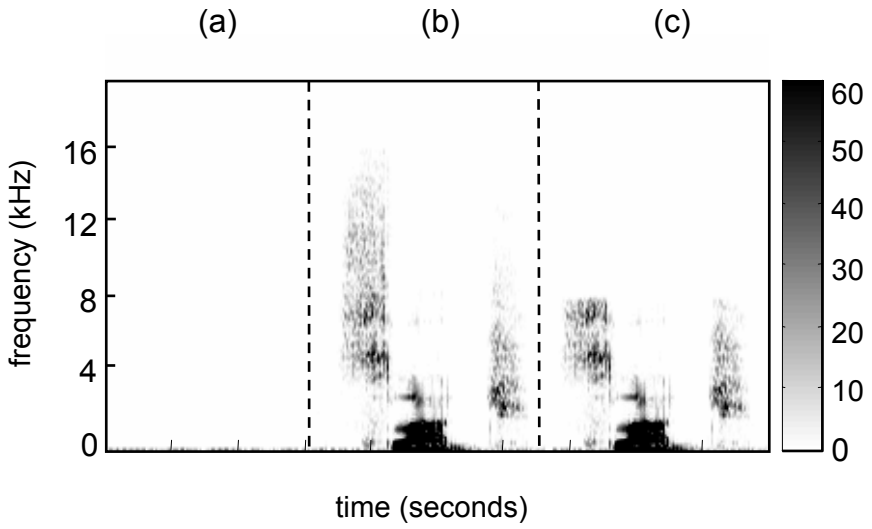



**Figure 5.2** Spectrograms of (a) a silent period, (b) the recorded word "sludge", and (c) the same word after low-pass filtering at 8 kHz. The spectrograms show that the recordings were made with high signal-to-noise ratio and that substantial energy exists above 8 kHz in the recorded speech.

## 5.2 Previous studies of speech localisation

A few studies have examined speech localisation in the horizontal plane using single words presented in virtual auditory space and found that although the lateral angle (the angle away from the median plane) was estimated as accurately as for non-speech broadband stimuli, there was an increase in front-back confusions (Begault and Wenzel, 1993; Ricard and Meirs, 1994). Using a larger range of locations, Gilkey and Anderson (1995) compared the localisation of click trains to recorded single words. They examined positions distributed randomly on a sphere (including elevations between -45° and 90°). They reported that performance was comparable in the left-right dimension, but poorer in the front-back dimension and in elevation for speech stimuli. The kinds of errors reported for speech stimuli are typical cone of confusion errors (see section 1.3.3), arising commonly in situations where the spectral cues to sound source location are somehow compromised.

In the present study, a similar approach to that of Gilkey and Anderson (1995) was used to examine speech localisation at a large range of locations on the sphere of space. However, in that study, subjects responded to their perceived locations on the sphere by pointing with a stylus to the analogous location on a spherical model that

was located in front of them. Importantly, the subjects were positioned using a bite-bar and could not see the model. Hence responses were guided only by feeling the contours that were etched onto its surface. In the present study, a nose-pointing technique was instead adopted, which has been argued to require less sensory translation and be more accurate (see Carlile, 1996a for review)

## 5.3 Experimental methods

### 5.3.1 Subjects and task

Five subjects (S1, S3, S4, S5, S7) participated in the experiments and all had previous experience in auditory localisation experiments. Experiments were carried out in the anechoic chamber and stimuli were presented in virtual auditory space (Chapter 2).

There were two experiments each consisting of three stimulus conditions. Experiment 1, which was completed by each subject before the commencement of Experiment 2, compared the ability of subjects to localise broadband noise, broadband speech and 8 kHz low-pass filtered speech. In Experiment 2, subjects localised speech stimuli in which the *level* of the high-frequency information was systematically varied. Each of the two experiments consisted of 1140 localisation trials in total, divided into 15 localisation tests. Thus in total, each subject completed 30 tests (approximately 8 hours of listening time) over a period of about two months.

The localisation tests were of the format described in section 2.4.1. Subjects were positioned in the centre of the chamber, with their heads calibrated to be facing directly ahead. Each test consisted of 76 trials in which the subject was presented with a stimulus and required to give a localisation response by pointing with their nose. An electromagnetic head-tracker recorded their estimates after a response button was pressed. No specific training was carried out for this experiment, as all subjects had been trained to localise in this set-up previously (section 2.4.2).

### 5.3.2 Speech stimuli

As a publicly available broadband speech corpus (i.e., not filtered at 8 kHz or below) could not be found, the Harvard list (Egan, 1948) was recorded in an anechoic environment by an actor with extensive vocal training. This corpus consists of 20 phonetically balanced word lists each containing 50 monosyllable words. These were recorded using a Brüel and Kjær 4165 microphone and a Brüel and Kjær 2610 amplifier. The speech was digitised at a sample rate of 80 kHz using an anti-aliasing filter with a 30 kHz cut-off. The duration of the recorded words ranged between 418 and 1005 ms with an average duration of 710 ms.

Because this experiment examines the influence of high-frequency spectral information on localisation, it was very important that background noise be eliminated. This was accomplished by: (i) seating the actor (in a quiet chair) close to the microphone; (ii) setting the amplifier so that signals were within and spanned the maximum range of the analogue to digital converter; (iii) turning off unnecessary equipment. Ultimately, the final test was that silent periods played back using VAS should sound silent. Figure 5.2 shows a spectrogram of a silent period and a recorded speech signal (the word “sludge”) under different filtering conditions. The spectrograms show that the recordings were made with high signal-to-noise ratio and that background noise was not a problem.

In total, five responses were obtained for 76 locations on the virtual sphere of space for each stimulus condition. For each of the five replicates, the spoken word signals were chosen randomly from a different balanced list of 50, thus five word lists were used for each experiment. For each condition, the same word stimuli were played from the same location on the sphere, thus keeping the stimulus set identical between conditions. Within each experiment, trials from the three conditions were interleaved and stimuli were presented in a randomised fashion.

### 5.3.3 Data analysis

In order to gauge the overall performance of subjects in the different experimental conditions, the spherical correlation coefficient (SCC, see section 2.4.3) was calculated. To investigate the pattern of responses more closely, the localisation data

were analysed in terms of the correspondence of response lateral and polar angles with target lateral and polar angles. This involved plotting actual target lateral (or polar) angle against perceived lateral (or polar) angle. Furthermore, lateral and polar angle errors were calculated for each trial and pooled in order to compare performance across stimulus conditions. Finally, in order to compare results to those of previous studies, it was useful to calculate the percentage of cone of confusion errors made by subjects in each condition. These were defined as large polar angle errors (> 90º in magnitude) and as such included the commonly reported front-back and up-down confusions.

## 5.4 Experiment 1: Speech localisation

### 5.4.1 Conditions

Experiment 1 consisted of three stimulus conditions. In condition A, which acted as a control, the stimulus was a 150 ms white noise burst that was windowed by applying a raised cosine to the first and last 10 ms. In condition B, stimuli were broadband speech signals, band-passed between 300 Hz and 16 kHz. Importantly, the duration of noise and speech stimuli were not matched, however if this is to have an effect, it would be expected that the longer duration of the speech stimuli might confer an advantage for localisation over the control broadband noise condition. Note also that as the stimuli were presented in virtual auditory space, any head movements made during stimulus presentation would not have provided any dynamic localisation cues. In condition C, stimuli were speech signals low-pass filtered at 8 kHz before DTF filtering. Low-pass filtering was performed in the time domain using brick-wall FIR filters (60 dB down in the stop band). The effect of the low-pass filtering is demonstrated in Figure 5.2 where it can be seen that certain parts of speech signals contain substantial high-frequency information that is removed with the 8 kHz low-pass filtering.

## 5.4.2 Results

### Overall performance

Table 5.1 shows the SCC calculated for each subject under the three stimulus conditions. The SCC was highest in the broadband noise condition (mean 0.88), fell slightly in the broadband speech condition (mean 0.85), and fell considerably in the low-pass speech condition (mean 0.65).

**Table 5.1** Spherical correlation coefficients (SCCs) for each of the five subjects in Experiment 1. Each of the three rows contains values for the three stimulus conditions. Each SCC is calculated on the basis of five repetitions at each of 76 stimulus locations (i.e. 380 trials in total). See section 2.4.3 for details of this statistic.

| | S1 | S3 | S4 | S5 | S7 |
|---|---|---|---|---|---|
| BB noise | 0.82 | 0.88 | 0.91 | 0.90 | 0.91 |
| BB speech | 0.75 | 0.87 | 0.87 | 0.88 | 0.87 |
| LP speech | 0.40 | 0.79 | 0.66 | 0.76 | 0.66 |

### Lateral and polar angle analysis

The correspondence between target and response lateral angles is illustrated in Figure 5.3. Each panel shows one stimulus condition (left: broadband noise; middle: broadband speech; right: low-passed speech). As all subjects showed a very similar pattern, their data were combined for plotting. It can be seen that response direction lateral angles correspond well to the target direction lateral angles in all conditions, as most of the data falls on or near the 'perfect response' diagonal.

As the polar angle data showed patterns that were highly individualised, the polar angle analysis is presented for each of the five subjects separately. Figure 5.4 illustrates the correspondence between target and response polar angles. Each row shows one subject, and again the three panels in a row represent the three stimulus conditions (left: broadband noise; middle: broadband speech; right: low-passed speech).

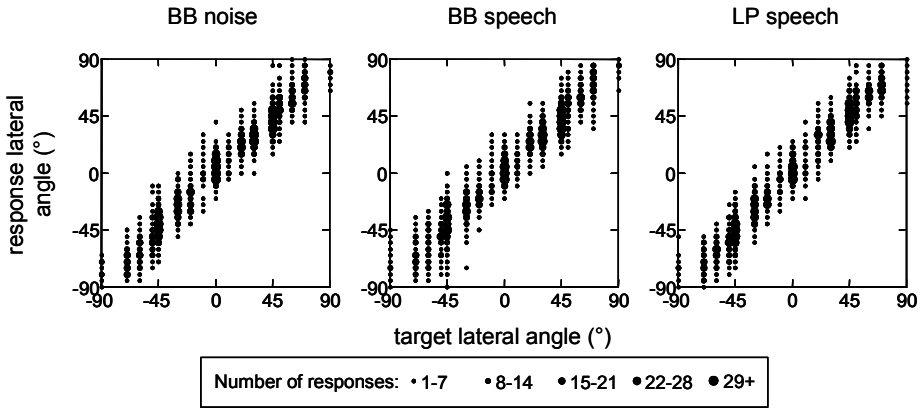



**Figure 5.3** Scatter plots showing lateral angle data (pooled across all subjects) for Experiment 1. The three panels contain data for the three stimulus conditions: broadband noise (left), broadband speech (middle) and low-pass filtered speech (right). Target lateral angle (abscissa) is plotted against response lateral angle (ordiante) and the size of the dots represents the number of responses clustered at a point.

It must be noted that these data are collapsed across lateral angle, and thus variations in the scale of polar angles are ignored. Referring to Figure 2.1b, it is clear that larger lateral angles (approaching ±90°) give rise to smaller polar angle circles on the co-ordinate sphere. This means that polar angle space becomes increasingly compressed at the sides, and thus small absolute errors can translate to extremely large polar angle errors. Thus in this kind of analysis one can expect to see many large polar angle errors overall, even in optimal listening conditions.

In the broadband noise control condition, all subjects estimated the polar angle relatively accurately, although S1 and S7 were slightly less accurate compared with the other subjects. In the broadband speech condition, performance was also relatively good. However all subjects showed some increase in error, which appears to be predominantly due to front-back confusions.

Individualised polar angle response patterns were most evident in the low-pass speech condition. S1 and S7 were dramatically affected by the low-pass filtering, showing a strong tendency to localise stimuli presented in the lower hemisphere of space (polar angle range -180º – 0º) to the upper hemisphere (polar angle range 0º - 180º). S3 and S5's performance was the most robust to low-pass filtering, but showed

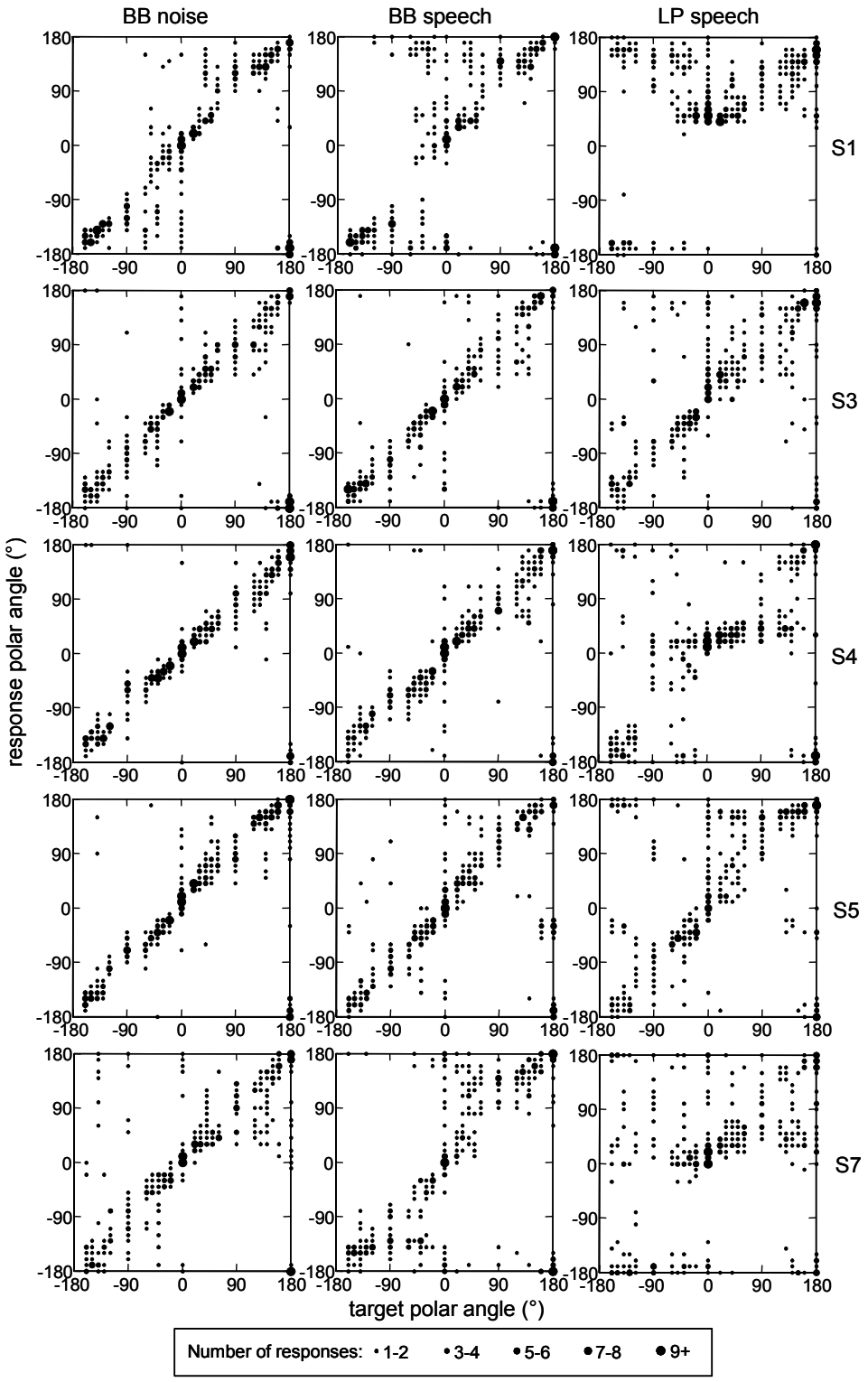



**Figure 5.4** Scatter plots showing polar angle data for Experiment 1. Each row shows data for a different subject as labelled. The three panels contain data for the three stimulus conditions: broadband noise (left), broadband speech (middle) and low-pass filtered speech (right). Target polar angle (abscissa) is plotted against response polar angle (ordiante) and the size of the dots represents the number of responses clustered at a point.

a spread in responses and some back-to-front confusions. Finally, S4 demonstrated some large polar angle response errors, particularly for stimuli presented in the lower frontal regions (polar angle range -90º - 0º).

In order to summarise the magnitude of the lateral and polar angle errors, the absolute value of the errors was calculated and the mean was obtained for each subject in each stimulus condition. Mean errors and standard errors of the means (SEMs) are shown in Figure 5.5. Lateral angle errors (Figure 5.5a) were small in general and did not appear to vary systematically across the three stimulus conditions (means of 8.3°, 8.9°, and 8.5°). A Kruskal-Wallis non-parametric ANOVA performed

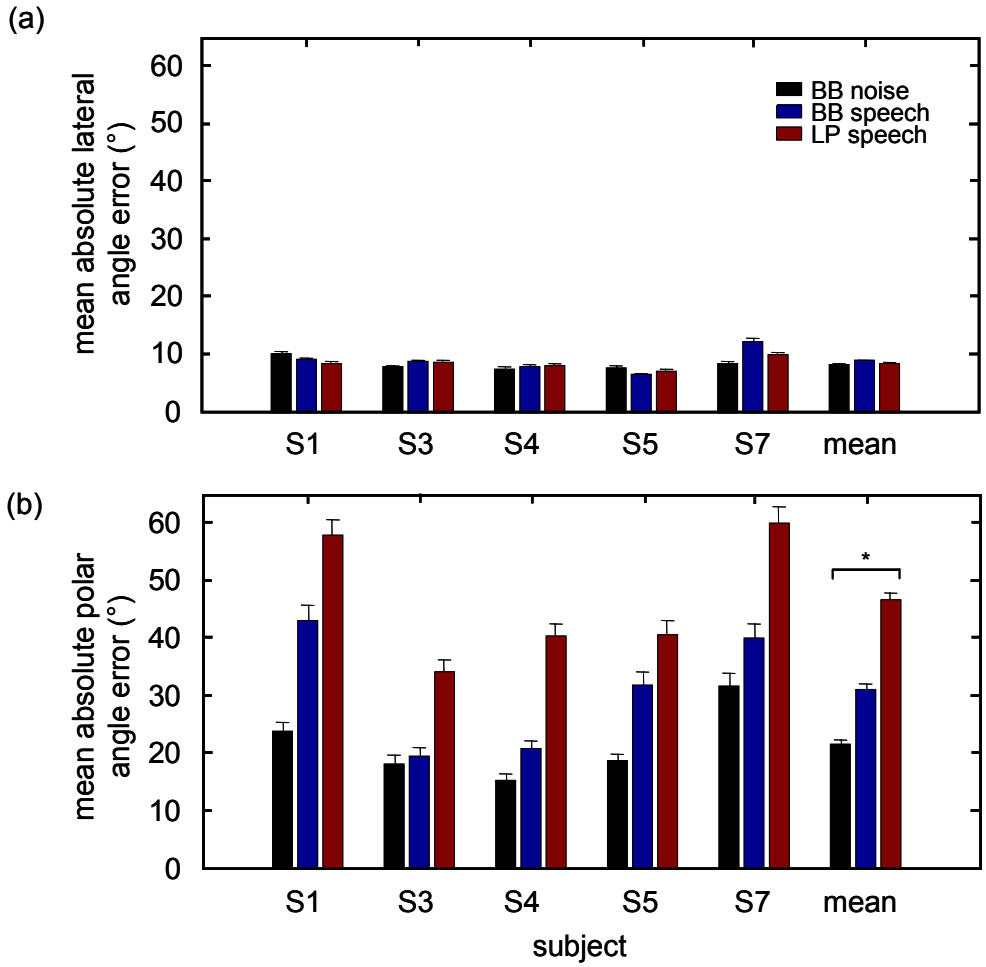



**Figure 5.5** (a) Mean absolute lateral angle errors from Experiment 1. The six clusters of bars show results for the five subjects as well as the mean across subjects. Mean errors are shown for broadband noise (black bars), broadband speech (blue bars) and low-pass filtered speech (red bars). Error bars show standard error of the mean. (b) Mean absolute polar angle errors from Experiment 1. All other details as for part a. The asterisk indicates that for the mean data all three conditions were significantly different from each other ($p<0.05$).

on the group data revealed no significant differences across conditions [$\chi^2(2,5697) = 5.35$, $p = 0.069$]. Mean polar angle errors (Figure 5.5b), however, varied considerably between stimulus conditions. The observation made above that S1 and S7 were most affected by the low-pass filtering is consolidated: their errors for that condition are larger than those of the other subjects. Overall, polar angle errors were consistently smallest in the broadband noise condition (mean 12.3°), higher in the broadband speech condition (mean 30.9°), and highest for low-passed speech (mean 46.4°). A Kruskal-Wallis non-parametric ANOVA performed on the group data revealed a highly significant effect of condition [$\chi^2(2,5622) = 401.94$, $p < 0.001$], and post-hoc analysis (Tukey HSD, $p = 0.05$) found significant differences between all three conditions.

Table 5.2 shows the calculated cone of confusion error rates. For different subjects, the number and distribution of cone of confusion errors across the sound conditions varied, although there were fewer errors on average for the broadband noise condition (mean 3.1%), an increase in the broadband speech condition (mean 8.4%) and a much larger number of errors for the low-pass speech condition (mean 16.4%).

**Table 5.2** Percentage of cone of confusion (COC) errors made by each of the five subjects in Experiment 1. Each of the three rows contains values for the three stimulus conditions. Each value represents the percentage of trials (out of the total 380) in which a COC error was made. See section 2.4.3 for details of this statistic.

| | S1 | S3 | S4 | S5 | S7 |
|---|---|---|---|---|---|
| BB noise | 5.8 | 2.9 | 1.8 | 2.9 | 1.8 |
| BB speech | 18.2 | 4.2 | 3.7 | 12.4 | 3.7 |
| LP speech | 22.6 | 11.6 | 15.5 | 16.6 | 15.5 |

## 5.4.3 Discussion

### Overall performance

The findings of the current study were consistent with those of previous studies. Subjects accurately estimated the lateral angle of a source regardless of spectral content, as the interaural cues provide a robust cue. This has been reported previously

for both speech (Ricard and Meirs, 1994; Gilkey and Anderson, 1995) and non-speech stimuli (Gilkey and Anderson, 1995; Carlile *et al.*, 1997). Performance in the control condition confirmed the well-supported notion that broadband flat-spectrum sounds are well localised in polar angle (e.g. Carlile *et al.*, 1997; King and Oldfield, 1997 and see Chapter 2). In terms of the localisation errors that arise when the monaural spectral cues are ambiguous, the present study confirmed and extended the findings of previous researchers. Our basic finding was that polar angle errors increased for sound stimuli with time-varying spectra (speech) but particularly when high-frequency content was removed (low-passed speech).

It was interesting to find that subjects were relatively accurate at localising broadband speech, despite some increase in confusions compared to broadband noise. Speech has a spectrum that is non-flat, and also varies over time, and yet listeners showed a reasonable capacity for distinguishing spectral cues to location from spectral features of the source spectrum. The remarkable ability of the auditory system to extract spatial cues from certain stimuli with non-flat spectra was discussed in Chapter 1, with particular reference to 'scrambled spectrum' and 'rippled spectrum' stimuli (see section 1.4.1). In that discussion, source familiarity was identified as an important factor in this process. In an experiment by Plenge and Bruenschen (1971), subjects were presented with speech from five loudspeakers in the upper median vertical plane. The authors proposed that familiar sources (with familiar spectra) would be localised better than unfamiliar sources because the spectrum due to directional filtering could be calculated. Indeed localisation performance dropped from 90% for familiar voices to 50% for unfamiliar voices. In the current experiment, as the same talker was used on every trial, it is likely that his voice quickly became familiar to listeners. Poorer polar angle estimates might be expected if the talker was varied randomly from trial to trial such that the source spectrum could not be predicted.

The low-pass filtered speech stimuli were included in this study because speech stimuli in the large majority of previous work have been low-pass filtered at 8 kHz or below. The results presented here indicate that low-pass filtered speech stimuli are localised much less accurately than broadband speech stimuli. Out of the literature consulted only two localisation studies used a cut-off frequency higher than 8 kHz. One was the Gilkey and Anderson study (1995), where speech stimuli were band-pass filtered 400 Hz to 11 kHz. Due to the differences in analyses, it is difficult to compare

performance of subjects in that study to the current one, but it is clear that the amount of error seen in Gilkey and Anderson's data is larger than that seen for our broadband speech (300 Hz – 16 kHz) condition. This suggests that frequency bands as high as 11 - 16 kHz are useful for accurate sound localisation of speech.

In another study, consonant-vowel (CV) localisation was examined in the presence of diffuse background noise (Karlsen, 1999). In that study, CVs were recorded in an anechoic environment with a low-pass cut-off frequency of 10 kHz. Localisation was tested in the horizontal plane. The author reported a relatively low percentage of front-back errors which, by inspection of the data, appears to be approximately 15% of the total trials. This is greater than the proportion of cone of confusion errors calculated in the present study for broadband speech (8.4%), suggesting that frequencies between 10 and 16 kHz can benefit speech localisation. However, the value reported by Karlsen is comparable to that reported here for low-passed speech (16.4%) despite the differences in the cut-off frequency. This seems to suggest that frequencies between 8 and 10 kHz do not greatly reduce front-back confusion. Of course these comparisons must be made with caution, as the testing locations and criteria for defining cone of confusion errors differed between the present study and that of Karlsen.

**The contribution of high frequency information**

An interesting analysis included in Karlsen's work looked at the differences in localisability between different CVs (Karlsen, 1999). In other words, he was probing the specific components of human speech that are useful for localisation. He found that the vowels /u/ and /a/ are localised more poorly than /i/. He also measured response time of subjects and noted that they seemed most sure (fastest response time) of CV locations if the CV contained the /s/ consonant. However, a surprising finding was that the strong consonants /s/ and /t/, containing high-frequency energy, resulted in a large number of front-back errors.

It was of interest for the present experiment to take this approach and examine whether the specific frequency content of different words was related to the localisation performance. Specifically, it was suspected that words with substantial energy above 8 kHz would be well localised and performance on these words would be *most* affected by the low-pass filtering. To examine this, a spectrogram of each of the 250 words used was calculated using a routine from the Auditory Toolbox for

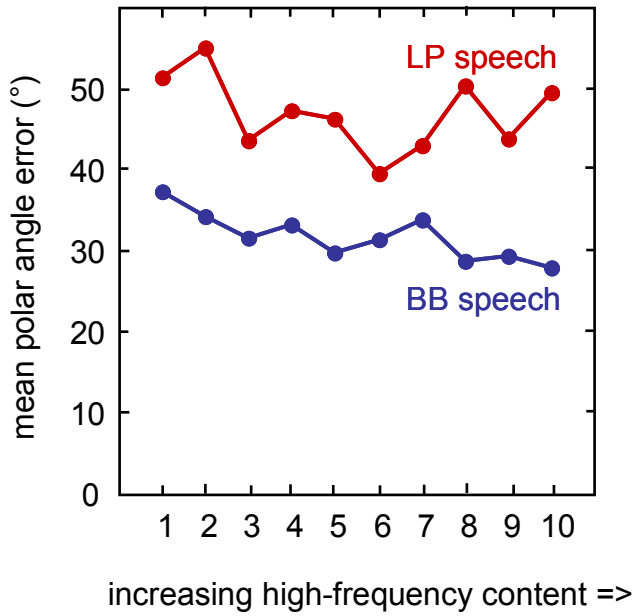



**Figure 5.6** Mean polar angle errors for different groups of words. The numbers 1 to 10 on the abscissa denote 10 groups of words of increasing high-frequency content (see section 5.4.3 for details). Mean errors are shown for broadband speech (blue) and low-pass filtered speech (red). It can be seen that *all* words were localised more poorly after low-pass filtering.

MATLAB (Malcolm Slaney, version 2, 1998). The total energy above 8 kHz was summed and taken as a metric (in arbitrary units) of high frequency content. The words were ranked on this basis, and then divided into 10 bins each containing 25 words. For each bin, polar angle errors were calculated for all trials (across all subjects) in which its members were the stimulus. Mean absolute polar angle errors were calculated independently for the broadband and low-pass filtered conditions.

Figure 5.6 shows the mean errors for the 10 bins. Note that the bins represent words with increasing amounts of high-frequency energy, i.e. bin 1 has the least and bin 10 has the most. Three main points emerge from this rather crude analysis. Firstly, it is clear that the curves do not change smoothly as a function of high-frequency energy. This is because the low-frequency content of these words is likely to vary somewhat independently of the high-frequency content, and the low-frequency energy will have an influence on the amount of error. Secondly, for the broadband condition, it can be seen that there is an overall decrease in the mean polar angle error with increasing high-frequency content. This suggests that words with more high frequency energy were better localised (about 10° error difference between bin 1 and bin 10). The low-passed condition shows a much more unstable curve, but since these

sounds were presented without any of their high-frequency energy, no relationship is expected here between high-frequency content and performance. Thirdly, for all bins, the mean errors are significantly lower in the broadband condition compared to the low-passed condition. Even for the first bin, i.e. the 25 words with the lowest amount of high-frequency energy, the improvement is substantial. Clearly these words have enough energy above 8 kHz to afford the auditory system a better estimate of the location.

To examine this point further, Figure 5.7a shows the power spectral density of the word with the *highest* calculated high frequency energy ("slap"). It can be seen that the energy in this signal remains high all the way up to 16 kHz (which was the band-pass cut-off) and a substantial amount of information is lost after low-pass filtering. For comparison, Figure 5.7b shows the power spectral density of the word with the *least* calculated high frequency content ("roe"). It can be seen that the energy drops off more quickly than for the previous example, but that there is indeed still significant energy above 8 kHz that is lost after low-pass filtering.

This analysis showed that all of the 250 word stimuli used in this experiment contained high-frequency information (above 8 kHz) that was useful for polar angle localisation. This emphasises that naturally spoken speech is a broadband stimulus, a notion that is frequently overlooked in the literature. Experiment 2 was conducted to test the hypothesis that the preservation of this high-frequency energy, even at a very low level, can benefit the localisation of natural speech.

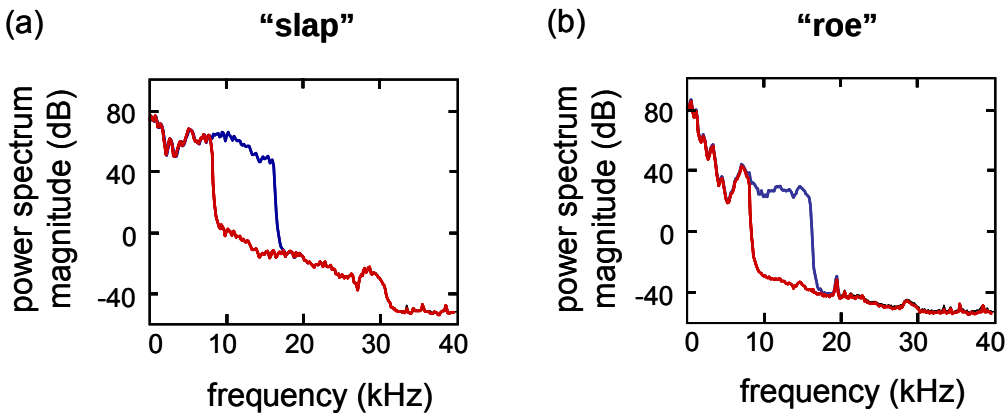



**Figure 5.7** (a) Power spectral density plots of the word "slap" with upper frequency cut-offs of 16 kHz (blue) and 8 kHz (red). (b) Power spectral density plots of the word "roe" with upper frequency cut-offs of 16 kHz (blue) and 8 kHz (red).

# 5.5 Experiment 2: Influence of high-frequency level

## 5.5.1 Conditions

Experiment 2 consisted of three stimulus conditions. Condition A acted as a control and stimuli were broadband speech signals (identical to those of Experiment 1 condition B). In conditions B and C, the level of the high-frequency information (above 8 kHz) was systematically varied. In condition B the high-frequency region was attenuated by 20 dB and in condition C it was attenuated by 40 dB. This was achieved by low-pass filtering the signals (as in Experiment 1 condition C) but with varying attenuation in the stop band. Figure 5.8 shows the power spectral density plots of a typical word under the different conditions. The low-pass condition from Experiment 1 is included for comparison as it was equivalent but with 60 dB attenuation in the stop band.

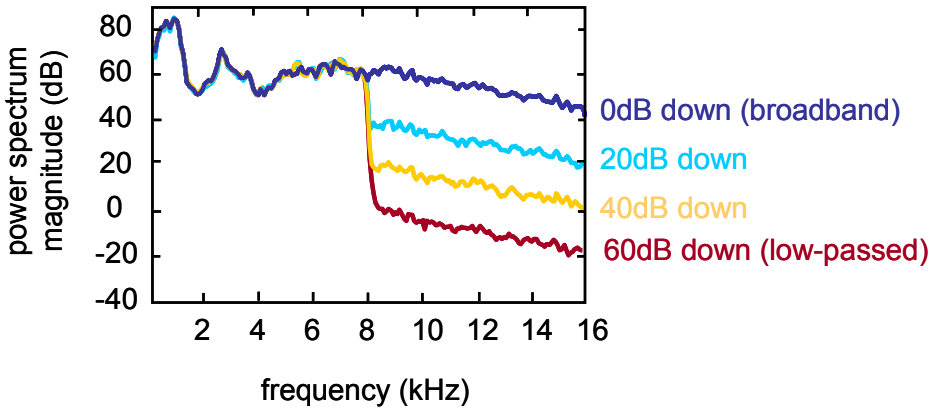



**Figure 5.8** An illustration of the stimulus conditions employed in Experiment 2. Shown are power spectral density plots of a speech stimulus after low-pass filtering with varying attenuation in the stop-band (0, 20, 40 and 60 dB down). Note that 60 dB down was not re-tested, but is equivalent to the low-pass speech condition of Experiment 1.

## 5.5.2 Results

### Overall performance

Table 5.3 shows the SCC calculated for each subject under the three stimulus conditions of this experiment as well as the low-pass condition of Experiment 1. The

SCC was highest in the broadband speech condition (mean 0.85) and in nearly all cases showed a gradual decline with increasing attenuation in the high-frequency region (means of 0.74, 0.70 and 0.65 for 20 dB, 40 dB and 60 dB down).

**Table 5.3** Spherical correlation coefficients (SCCs) for each of the five subjects in Experiment 2. The first three rows contain values for the three new stimulus conditions and the fourth row is a reiteration of the data from the low-pass condition of Experiment 1. Each SCC is calculated on the basis of five repetitions at each of 76 stimulus locations (i.e. 380 trials in total). See section 2.4.3 for details of this statistic.

| | S1 | S3 | S4 | S5 | S7 |
|---|---|---|---|---|---|
| BB speech | 0.78 | 0.88 | 0.86 | 0.87 | 0.86 |
| 20 dB down | 0.57 | 0.82 | 0.75 | 0.82 | 0.75 |
| 40 dB down | 0.56 | 0.77 | 0.70 | 0.78 | 0.70 |
| LP speech | 0.40 | 0.79 | 0.66 | 0.76 | 0.66 |

### Lateral and polar angle analysis

The correspondence between target and response lateral angles for the three speech conditions is illustrated in Figure 5.9 (left: broadband; middle: 20 dB down above 8 kHz; right: 40 dB down above 8 kHz). As all subjects showed a very similar pattern, their data were combined for plotting. It can be seen that target lateral angles correspond well with response lateral angles in all conditions, as most of the data falls on or near the 'perfect response' diagonal.

As in Experiment 1, the polar angle data were highly individualised. Figure 5.10 illustrates the correspondence between target and response polar angles. Each row shows one subject, and again the three panels in a row represent the three speech conditions (left: broadband; middle: 20 dB down above 8 kHz; right: 40 dB down above 8 kHz).

In the broadband speech condition, all subjects estimated the polar angle with good accuracy, as was seen in Experiment 1. Again, S1 and S7 were less accurate localisers, making more front-to-back confusions. When the high-frequency level was lowered, S1 and S7 were the most affected. They showed a dramatic increase in error for the 20 dB down condition, which was exacerbated in the 40 dB

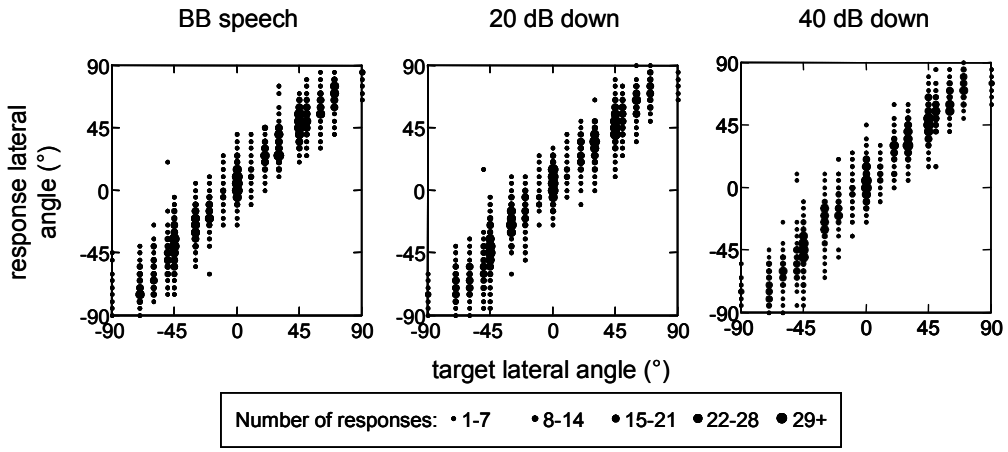



**Figure 5.9** Scatter plots showing lateral angle data (pooled across all subjects) for Experiment 2. The three panels contain data for the three stimulus conditions: broadband speech (left), 20dB down low-pass speech (middle) and 40 dB down low-pass speech (right). Target lateral angle (abscissa) is plotted against response lateral angle (ordiante) and the size of the dots represents the number of responses clustered at a point.

down condition. As was seen for low-passed speech in Experiment 1, these two subjects had strong tendency to localise stimuli presented in the lower hemisphere of space (polar angle range -180º – 0º) to the upper hemisphere (polar angle range 0º - 180º). The other subjects were less affected by the 20 dB drop in high-frequency level, showing just a small spread in responses. For the 40 dB down condition, this spread in error was more severe and many large errors are evident.

Figure 5.11 summarises the magnitude of the lateral and polar angle errors. Mean lateral and polar angle errors (with SEMs) are shown for each subject under the three conditions of this experiment. Data from the low-pass condition of Experiment 1 are also included for comparison. Mean lateral angle errors (Figure 5.13a) were again small and did not vary systematically across conditions (means of 9.4°, 9.7°, 9.6° for 0 dB, 20 dB, and 40 dB down). A Kruskal-Wallis non-parametric ANOVA performed on the group data revealed no significant differences across conditions [$\chi^2(2,5697) = 1.68$, $p = 0.432$]. Mean polar angle errors (Figure 5.13b), however, varied considerably across conditions. Values were consistently smallest in the control broadband speech condition (mean 26.6°) and increased with increasing attenuation in the high-frequency region (means of 38.0° and 42.3° for attenuations of 20 dB and 40 dB respectively). A Kruskal-Wallis non-parametric ANOVA performed on the group data revealed a highly significant effect of condition [$\chi^2(2,5622) = 128$, $p < 0.001$],

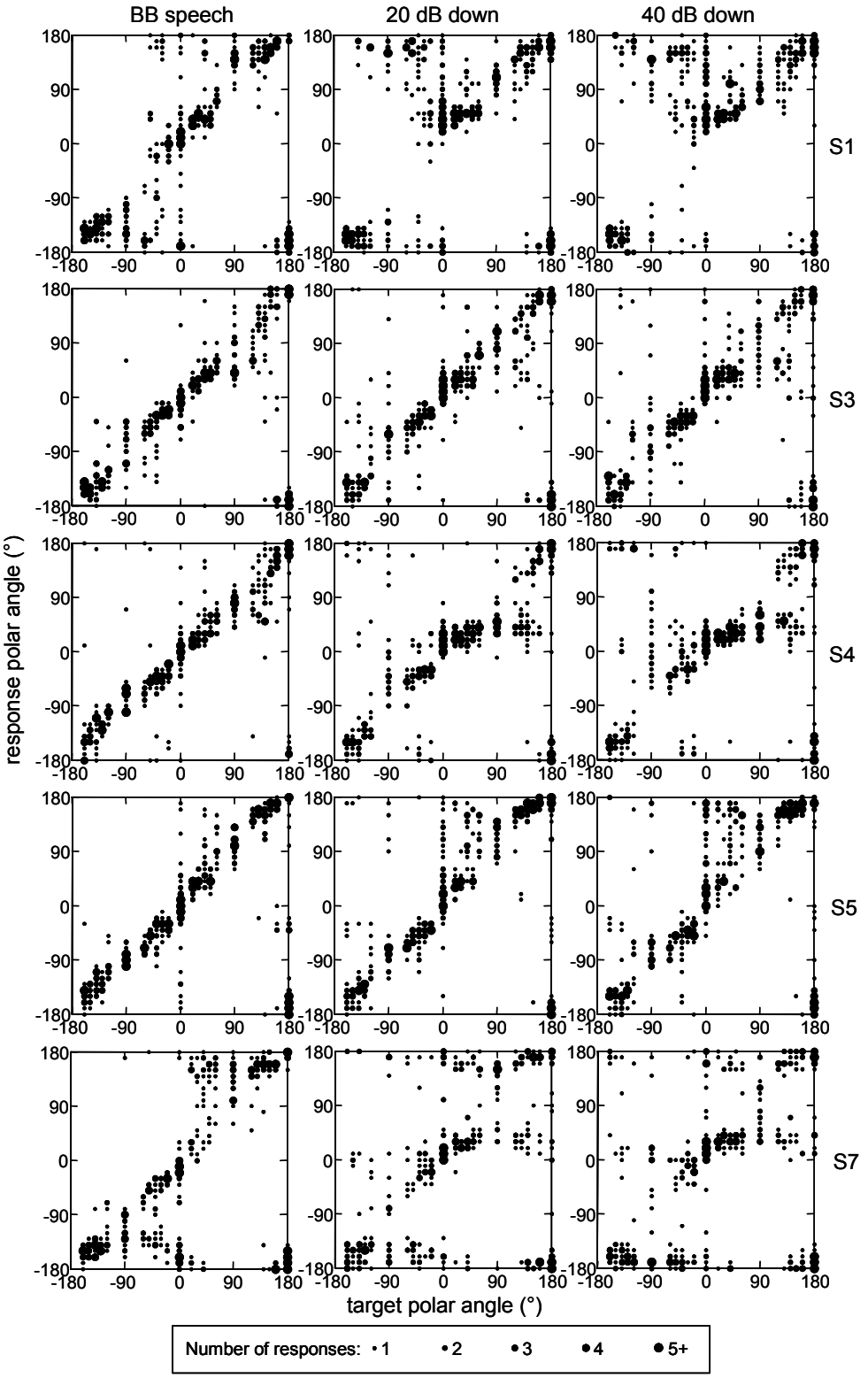



**Figure 5.10** Scatter plots showing polar angle data for Experiment 2. Each row shows data for a different subject as labelled. The three panels contain data for the three stimulus conditions: broadband speech (left), 20dB down low-pass speech (middle) and 40 dB down low-pass speech (right). Target polar angle (abscissa) is plotted against response polar angle (ordiante) and the size of the dots represents the number of responses clustered at a point.

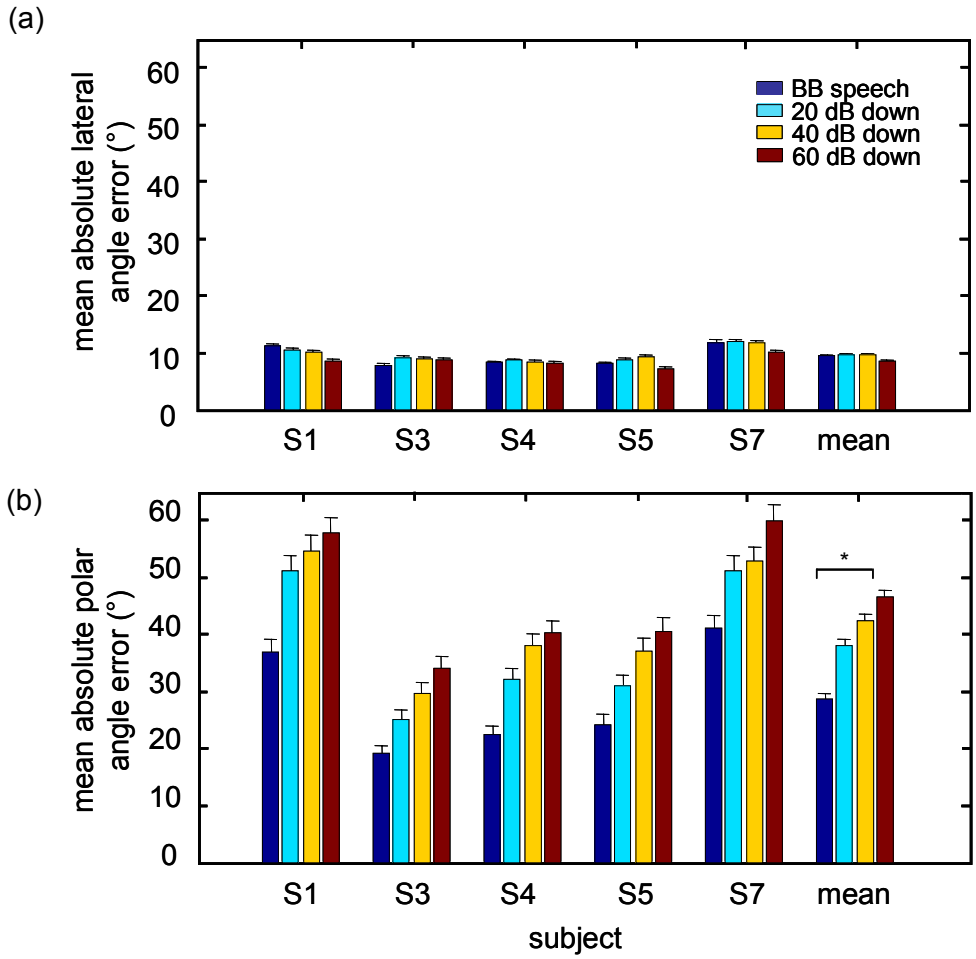



**Figure 5.11** (a) Mean absolute lateral angle errors from Experiment 2. The six clusters of bars show results for the five subjects as well as the mean across subjects. Mean errors are shown for broadband speech (dark blue bars), 20 dB down low-pass speech (light blue bars) and 40 dB down low-pass speech (yellow bars). Errors for low-pass filtered speech (60 dB down) from Experiment 1 are also shown for comparison (red bars). Error bars show standard error of the mean. (b) Mean absolute polar angle errors from Experiment 2. All other details as for part a. The asterisk indicates that for the mean data the three new conditions were significantly different from each other ($p<0.05$).

and post-hoc analysis (Tukey HSD, $p = 0.05$) found significant differences between all three conditions. Inspection of the low-pass data from Experiment 1 (60 dB down) suggests that this gradual trend continues with further decreases in high-frequency level. Furthermore, the observation made above, that S1 and S7 seemed to be most dramatically affected by even a small drop in high-frequency level, is confirmed.

Table 5.4 shows the calculated cone of confusion error rates. Individual differences are apparent, and most notably S1 was particularly prone to this kind of error in all of the low-passed conditions. On average, the error rate increased with

increasing attenuation in the high-frequency region (means of 6.7%, 11.9%, 13.8% and 16.4% for attenuations of 0 dB, 20 dB, 40 dB, and 60 dB respectively).

**Table 5.4** Percentage of cone of confusion (COC) errors made by each of the five subjects in Experiment 2. The first three rows contain values for the three new stimulus conditions and the fourth row is a reiteration of the data from the low-pass condition of Experiment 1. Each value represents the percentage of trials (out of the total 380) in which a COC error was made. See section 2.4.3 for details of this statistic.

| | S1 | S3 | S4 | S5 | S7 |
|---|---|---|---|---|---|
| BB speech | 13.4 | 3.4 | 5.3 | 6.1 | 5.3 |
| 20 dB down | 20.8 | 6.6 | 10.8 | 10.8 | 10.8 |
| 40 dB down | 23.4 | 6.6 | 12.6 | 13.9 | 12.6 |
| LP speech | 22.6 | 11.6 | 15.5 | 16.6 | 15.5 |

### 5.5.3 Discussion

These results showed that the polar angle localisation of speech is related to both the *presence* and the *level* of high-frequency information in the stimulus. There are two possibilities to explain the gradual decline in performance with increasing attenuation in the region above 8 kHz. One option is that more and more words in the corpus have their high-frequency content attenuated to an inaudible level (depending on the original level) and thus the error rate may be related to the *number* of broadband words presented. The extreme case would be the low-pass speech (60 dB down) condition, where it is assumed that no words carry high-frequency information. A second possibility is that most words retain some high-frequency content in the 20 and 40 dB down conditions, and that the system has an ability to make use of this low-level energy. Indeed the analysis in Experiment 1 (section 5.4.3) showed that words with very low high-frequency energy were localised better if this energy was preserved. If this is the case, then the gradual decline in performance with increasing attenuation must mean that localisation is impaired at low levels.

It has been observed using non-speech stimuli that localisation accuracy is a function of stimulus level (Harris, 1998). Using broadband noise, it was reported that elevation perception worsened and front-back confusions increased if the stimuli were

presented at low sound pressure levels (close to the audibility threshold). Furthermore, Abouchacra *et al.* (1998) presented speech phrases in diffuse noise, and reported that localisation accuracy in the horizontal plane improved as signal-to-noise ratio increased (18%, 89% and 95% accuracy at –18 dB, +12 dB and in quiet respectively). Although these level manipulations were not restricted to the high-frequency region, they demonstrate that the auditory localisation system has an ability to utilise localisation cues that appears to depend on the level of the signal.

## 5.6 General discussion

One of the striking features of the data collected in these experiments is the presence of strong individual differences. The subjects varied in their baseline localisation accuracy, their vulnerability to low-pass filtering of the speech stimuli, and their error patterns. A likely explanation for the differences seen with low-pass filtering is that some subjects are more reliant on high-frequency spectral features for localisation than others. As the high-frequency spectral cues derive from physical outer ear features, subjects with smaller ears (conchae in particular) will produce spectral cues at higher frequencies than subjects with larger ears. This will mean that they are more affected by low-pass cut-offs in the higher frequency regions. Indeed the two subjects who were most affected by the 8 kHz low-pass filtering in these experiments (S1 and S7) were the only two females in the group and had smaller ears than the males.

In a surprising result, Begault *et al*. (2001) reported that the use of individualised HRTFs was of no benefit for the localisation of natural speech signals (3 second samples of conversational speech) presented in VAS. Unfortunately the authors did not specify what the upper frequency cut-off of their speech stimuli was. However, they argued that individualisation did not have an effect because most of the spectral energy of speech is in a frequency region where ITD cues are more significant than spectral cues. While this may be true, the current experiment has shown clearly that the high-frequency energy in speech can and does aid in localisation. There are several reasons why individualisation may not have affected localisation in the study of Begault and colleagues. Firstly, only six azimuth positions on the 0° elevation plane were examined. This corresponds to a relatively weak test, as responses would have been highly constrained by the expected locations. Secondly,

no locations off the 0° elevation plane were examined, and thus many locations where spectral cues are perhaps the most relevant were not considered.

One of the aims of the work presented in this chapter, and the work of others, was to define the features important in optimising the spatial perception of speech. It has been shown here that the inclusion of high-frequencies, which are so often filtered out of speech in playback situations, can greatly improve the spatial perception. The reason this is thought to be important is because spatial perception plays an important role in competing source situations. However, there is also some evidence that the preservation of high-frequency speech information can be of benefit in non-spatial tasks. For example, it has been shown that children and adults require an upper cut-off frequency of 9 kHz to optimally identify fricatives (such as the ”s” sound) in quiet (Stelmachowicz *et al.*, 2001). Furthermore, Vickers *et al.* (2001) examined the intelligibility of nonsense vowel-consonant-vowel sounds presented monaurally in quiet, and systematically varied the low-pass cut-off of the speech sounds. They showed (for listeners with mild high-frequency hearing loss) that intelligibility increased steadily with increasing cut-off frequency (800 Hz to 8 kHz), demonstrating that the inclusion of higher frequencies is beneficial for intelligibility. While this study did not employ higher cut-off frequencies, it can be seen in their data that for five out of six ears, performance had not reached a maximum level in the 8 kHz condition.

## 5.7 Conclusions

A broadband speech corpus (300 Hz – 16 kHz) was used to investigate the ability of human listeners to localise monosyllabic words. Experiment 1 showed that low-pass filtering the stimuli at 8 kHz dramatically degraded performance and increased errors associated with the cone of confusion. Experiment 2 showed that the preservation of information above 8 kHz, even at a low level, provided a benefit for polar angle localisation. Although the lower frequencies (below 8 kHz) are known to be sufficient for accurate speech recognition in most situations, these results demonstrated that natural speech contains information between 8 and 16 kHz that is essential for accurate spatial perception.

# Chapter 6: Spatial performance with concurrent speech sources

## 6.1 Introduction

It was clear from the findings of Chapter 5 that speech is a broadband stimulus, and it is localised well by human listeners when its high frequencies are retained. It was then possible to repeat the experiments presented in Chapters 3 and 4 (which used broadband non-speech stimuli) using broadband speech stimuli. This allowed an examination of auditory spatial resolution and spatial interactions with a more natural stimulus of high salience in human environments.

Very few studies were found in the literature that dealt with the spatial perception of simultaneous speech sources. Those that were found presented competing sentences in the frontal horizontal plane only. Drullman and Bronkhorst (2000) used low-pass filtered speech (4 kHz cut-off) presented in virtual auditory space and subjects had to recognise and localise a target talker in the presence of 1-4 competitors. Performance was reported to be poor (around 50% correct on average) and was slightly worse with more competitors. However, using a similar set-up, Yost *et al.* (1996) asked subjects to localise all of the sources, and allowed an unlimited number of presentations of the stimulus. Good levels of performance were reported in this study for up to three talkers. Hawley and colleagues also reported high levels of performance when one out of three talkers had to be localised (Hawley *et al.*, 1999).

In terms of variations in spatial resolution, the studies using speech low-pass filtered at 4 kHz both reported that performance was slightly better at the front in comparison to the sides (Drullman and Bronkhorst, 2000; Yost *et al.*, 1996). However, the study that employed broadband speech reported no effect of stimulus configuration (Hawley *et al.*, 1999). No studies could be found that examined the localisation of speech in a competing source environment for locations off the horizontal plane. Furthermore, no studies were found that systematically examined

spatial discrimination using speech stimuli. The two experiments described in this chapter were carried out to address these matters.

In Experiment 1, the two-point discrimination experiment described in Chapter 3 was modified to measure spatial resolution for simultaneous speech stimuli in the horizontal and vertical dimensions. Instead of the two-point discrimination approach, where subjects had to judge the number of sources, a direction discrimination approach was chosen where subjects had to explicitly judge the relative location of two simultaneous stimuli.

In Experiment 2, absolute speech localisation was again examined, but in the presence of a simultaneous speech source. An identical paradigm to that described in Chapter 4 was adopted. The goals were to (a) observe whether the accuracy of speech localisation seen in Chapter 5 would hold up in the presence of a competing source, and (b) determine whether spatial interactions occur with concurrent speech sources such as those observed with non-speech sources in Chapter 4.

## 6.2 Experiment 1: Location discrimination with paired speech sources

### 6.2.1 Experimental methods

**Subjects and task**

This experiment was essentially an extension of the auditory two-point discrimination experiments described in Chapter 3. Three subjects participated (S1, S3 and S6) and two of these were participants in the earlier two-point discrimination experiments (S1 and S3). The experiment was carried out in the soundproof room and stimuli were presented in virtual auditory space (Chapter 2).

The experiment consisted of 10 tests, each of which took around 15 minutes to complete. Subjects completed the 10 tests one or two at a time over a period of about three weeks. Each test consisted of 180 trials in which the subject was presented with a simultaneous pair of stimuli (a target and a masker). In five of the tests, stimuli were separated horizontally and in the other five tests the stimuli were separated vertically. The task was a spatial discrimination one and took the form of two-alternative forced choice. Subjects were required to indicate, by pressing one of two buttons on a hand-

held response box, whether the target was located to the left or the right of (or above or below) the masker. The two source locations were termed the 'reference' and the 'test' locations as described in the next section.

**Stimuli**

For each of the five tests in the horizontal condition, a target word was chosen randomly from the corpus of monosyllabic words described in Chapter 5 (section 5.3.2). This word was identified to the subject at the beginning of the test, and it served as the target for the 180 trials of the test. The masking word was kept constant throughout the test, and was chosen from the corpus such that it matched the target word as closely as possible in duration (see Figure 6.1 for an illustration). The same five word pairs were used for the five vertical discrimination tests.

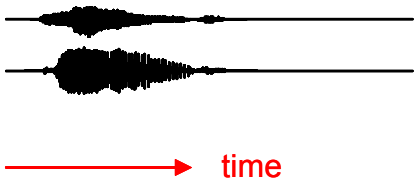



**Figure 6.1** An example of a stimulus pair used in Experiment 1. Spoken words were paired such that they matched closely in duration.

The generation of target/masker stimulus pairs followed essentially the same steps as those described in the previous chapters. Stimuli were spatialised using individualised DTFs (Chapter 2). In order to present the target and masker simultaneously, the two binaural stimuli were generated independently and added. The two words were of approximately equal level, giving an overall signal-to-noise ratio of approximately 0 dB. After summing, stimuli produced a sensation level of approximately 50 dB.

Two band-pass conditions were employed during this experiment. During half of the trials, the speech stimuli were broadband (300 Hz – 16 kHz) and for the other half of the trials, low-pass filtered (300 Hz – 8 kHz). These filtering conditions were the same as those used in the speech localisation experiment (Chapter 5) and the filtering technique was identical.

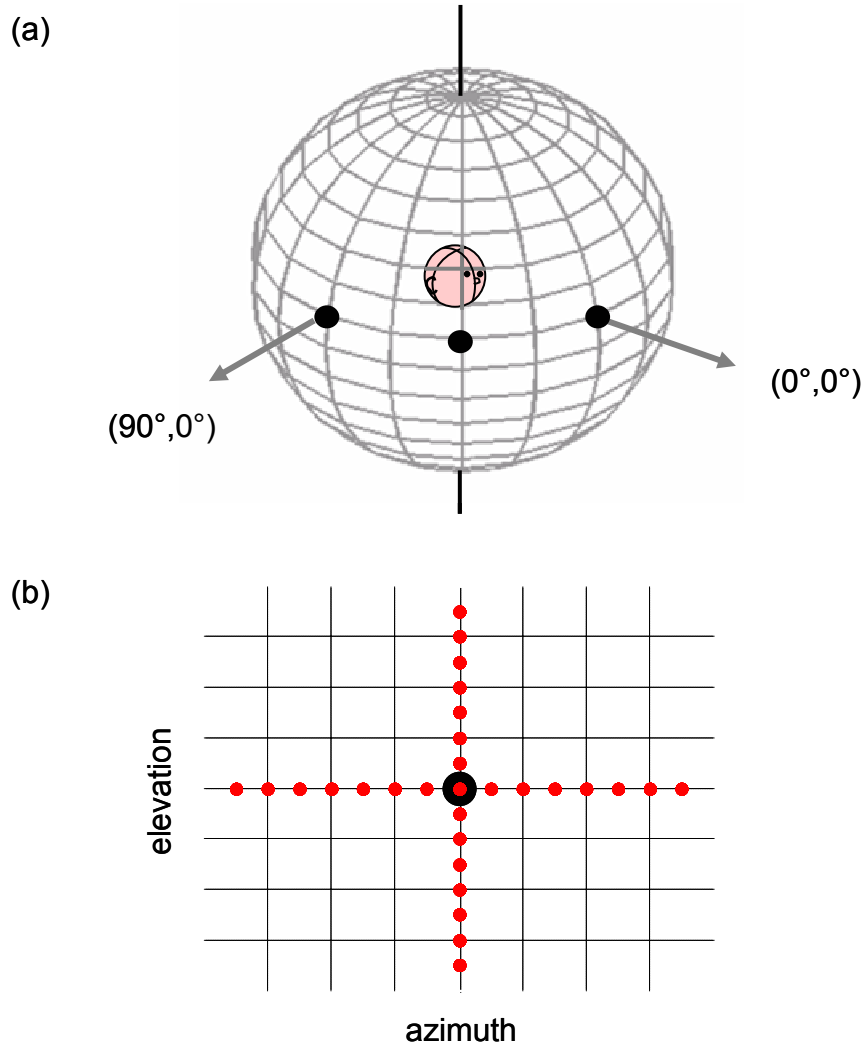



**Figure 6.2** Stimulus configurations, described using the azimuth/elevation co-ordinate system. (a) Three reference locations were employed in the right frontal quadrant on the 0° elevation plane: azimuths 0°, 45°, and 90° (shown as black dots). (b) To create stimulus location pairs, each of these reference locations was paired with 15 test locations distributed in azimuth and elevation (red dots, see section 6.2.1 for details).

### Stimulus configurations

Stimulus configurations were similar to those described in Chapter 3 for the two-point discrimination experiment, but a reduced set of reference locations was used. Reference locations of 0°, 45° and 90° were used (Figure 6.2a), and test locations were distributed symmetrically in azimuth and elevation about the reference location (Figure 6.2b). Note that this experiment is based on a single-pole co-ordinate system (section 2.2) and thus changes in elevation at the lateral azimuths (45° and 90°) cross over cones of confusion. For each reference location, the same 15 test locations were

used as those described in Chapter 3. For vertical separation, the testing range was the same for each of the three locations, and spanned the entire available range of elevations: from -45° to 90° (directly overhead). Ranges for horizontal separation varied with location as required to cover a suitable range: ± 21° for 0° azimuth; ±42° for 45° azimuth; ±63° for 90° azimuth. Importantly, the positive separation values at reference location 90° resulted in some stimuli occurring behind the interaural axis (azimuths>90°). Subjects were instructed that such 'clockwise' stimuli were to be judged as 'to the right' of 90°. The 15 test locations were paired with each of the three reference locations under the two bandwidth conditions in a single test. These 90 combinations were presented twice in a random fashion, once with the target at the reference location and once with the masker at the reference location. Thus there were two replicates obtained for each stimulus arrangement in a single test. As there were five tests, ten replicates were collected in total for analysis. These were used to plot psychometric curves similar to those described in Chapter 3.

### 6.2.2 Results

Psychophysical curves for horizontal discrimination are shown in Figure 6.3. Data from the individual subjects are shown in (a) to (c) and mean data are shown in (d). The three subplots in a panel show data for each of the three reference locations as labelled. The abscissa shows the separation of the test stimulus with respect to the reference, where negative values indicate leftward separation and positive values indicate rightward separation. Plotted on the ordinate is the percentage of trials (out of 10) where the subject reported that the stimulus at the test location was to the right of that at the reference location. Results are shown for broadband speech (solid lines) and low-pass filtered speech (dotted lines).

All subjects displayed the best discrimination at 0° azimuth, as indicated by the steep curves, and the bandwidth appeared to have minimal influence on the results. The shallower curves at 45° indicate that resolution was coarser at this location. For S3 and S6, low-pass filtering caused some disturbance at the more lateral test positions as indicated by the troughs in the curves. Discrimination was poorer again at 90°, and each subject displayed a different pattern of results. S1 performed reasonably well in the broadband condition, and was able to judge stimuli that were displaced

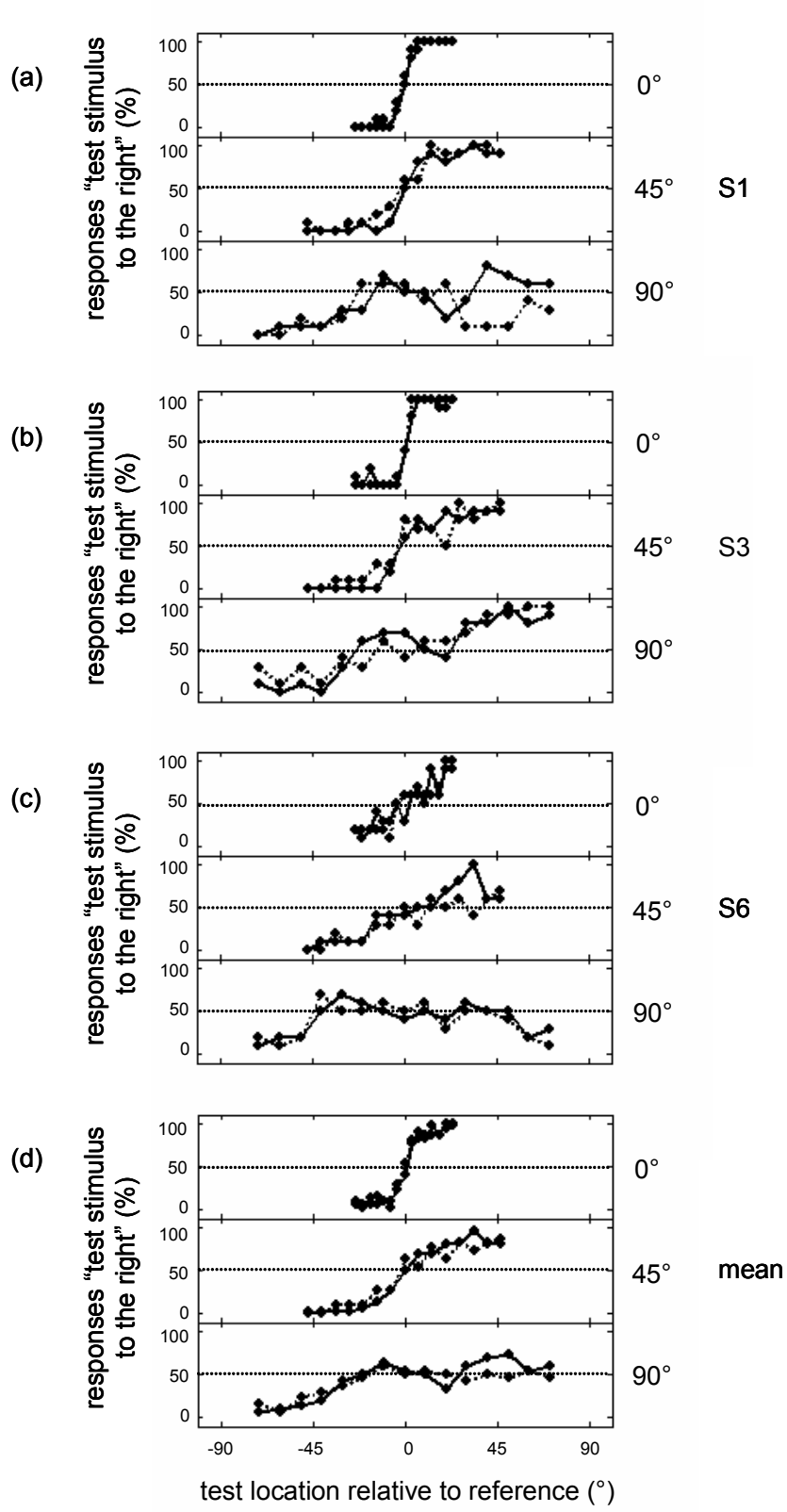



**Figure 6.3** Psychophysical curves for horizontal separation in Experiment 1. (a) to (c) show data for subjects S1, S3 and S6 respectively and (d) is mean data. The three subplots in each panel show data for each of the three reference locations (top to bottom: 0°, 45°, and 90° azimuth). For the given test locations (abscissa), the curves show the percentage of trials in which the subject responded that the test location was to the right of the reference location. Results are shown for broadband speech (solid lines) and low-pass filtered speech (dotted lines).

behind the interaural axis. However, with low-pass filtering this ability was lost and responses were generally reversed. It appears that S1 reflected posterior stimuli to the front in this situation. S3 was capable of correct horizontal discrimination at 90° in both broadband and low-pass conditions, even when front/back discrimination was required. S6 was able to discriminate horizontally when the test location was towards the front, but appeared unable to maintain performance past the interaural axis, even in the broadband condition.

Psychophysical curves for vertical discrimination are shown in Figure 6.4. Details are as for Figure 6.3 except that on the abscissa, negative values indicate downward separation and positive values indicate upward separation. Plotted on the ordinate is the percentage of trials (out of 10) where the subject reported that the stimulus at the test location was *above* that at the reference location. Results are again shown for broadband speech (solid lines) and low-pass filtered speech (dotted lines).

In general, vertical discrimination was poor. The curves are all relatively flat, indicating that even with large separations of the two words subjects could not judge the relative location of the two. Taking each subject and location in turn, however, there are some variations.

At 0°, S1 and S6 could not discriminate vertically at all within the testing range, regardless of bandwidth, as the curves do not deviate greatly from 50% (chance performance). S3 appeared to have some ability to discriminate 0° elevation from lower elevations but not from higher ones (for both bandwidth conditions).

At 45°, S6 showed no improvement. S1 however showed some ability to discriminate when the test location was high in elevation (for both bandwidth conditions). S3 discriminated well at this location, performing slightly better for broadband stimuli.

At 90°, S1 was able to discriminate when the test location was elevated, but only for broadband stimuli. When the test location was down low, performance remained near chance. S3 had a reasonable ability to discriminate with broadband stimuli but performance was poorer for low-passed stimuli. The curves at 90° for S6 are erratic, and responses do not appear to be related to separation of the two stimuli.

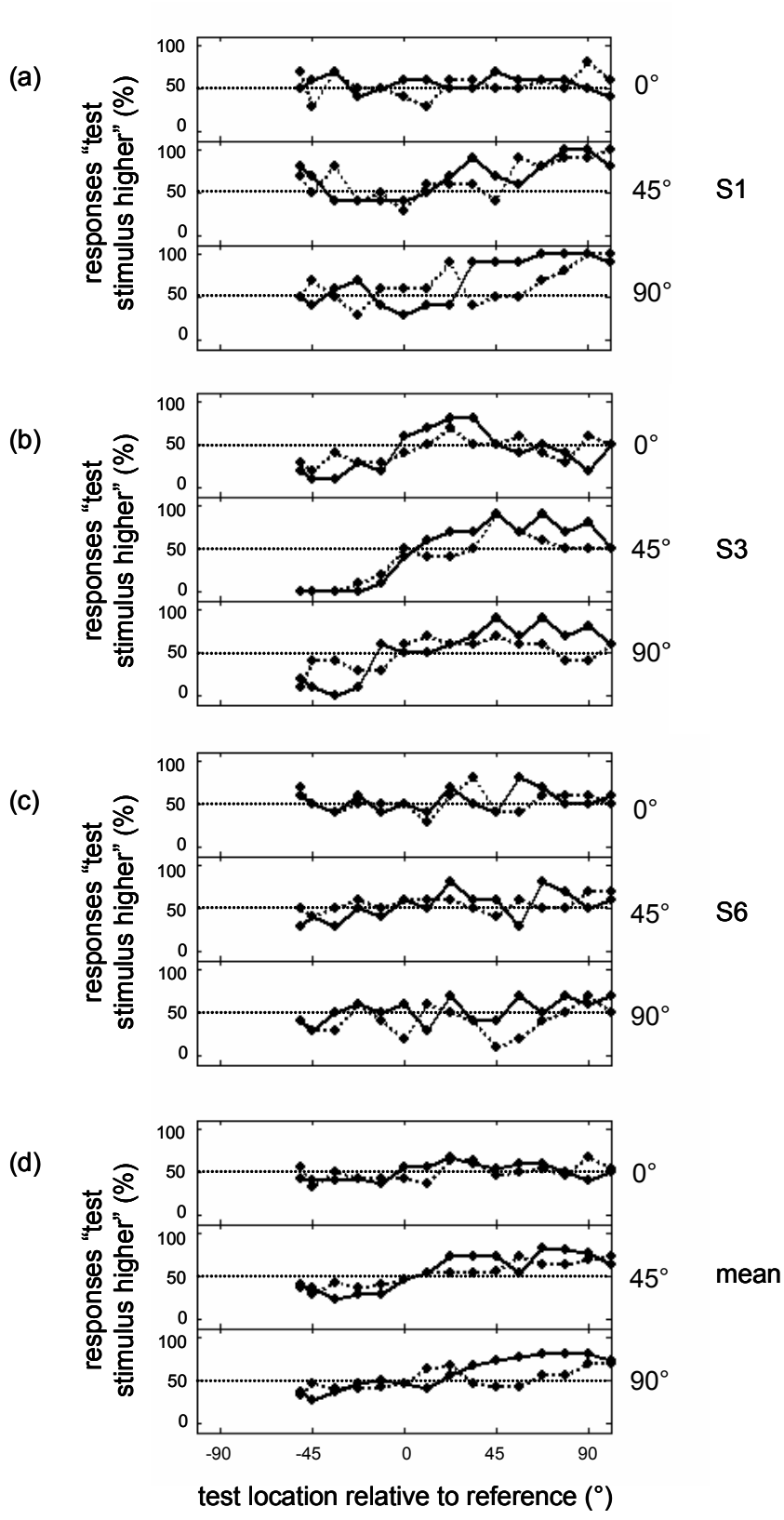



**Figure 6.4** Psychophysical curves for vertical separation in Experiment 1. (a) to (c) show data for subjects S1, S3 and S6 respectively and (d) is mean data. The three subplots in each panel show data for each of the three reference locations (top to bottom: 0°, 45°, and 90° azimuth). For the given test locations (abscissa), the curves show the percentage of trials in which the subject responded that the test location was above the reference location. Results are shown for broadband speech (solid lines) and low-pass filtered speech (dotted lines).

### 6.2.3 Discussion

**Thresholds for horizontal speech discrimination**

There was a clear trend in the horizontal broadband data showing that for increasing azimuth, a larger separation of the stimulus pair was required for correct and reliable left/right discrimination. In order to quantify this effect, discrimination thresholds were estimated from the psychophysical curves for all subjects. In order to compare these thresholds with the two-point discrimination thresholds estimated in Chapter 3 (section 3.5.3), threshold was defined as the separation value whereby the relative source locations were correctly reported at a rate 75% better than chance (i.e. a response rate of 87.5%). As 'positive' and 'negative' separation values were tested (corresponding to left and right, or up and down) a threshold was obtained separately from the two halves of each curve. It was not possible to obtain a value in some cases, where performance did not reach the 87.5% level within the range of testing.

Figure 6.5a, b, and c illustrate these angular thresholds for subjects S1, S3 and S6 respectively. The three groups of columns represent the three reference positions, and in each group there are either two, one, or no thresholds that could be obtained for a particular subject. Although there are bars missing, it can be seen that horizontal discrimination thresholds show an increase with increasing lateral position. Using the available individual thresholds, mean thresholds were calculated for each reference azimuth (Figure 6.5d). These values confirm the trend for increasing threshold with increasing reference azimuth, which is also consistent with the two-point discrimination results of Chapter 3 (section 3.5.3). Clearly there are similar constraints on the ability to resolve two source positions and the ability to judge relative location. To compare the magnitude of thresholds in these two experiments, the two-point discrimination thresholds at reference azimuths of 0°, 22.5°, 45°, 67.5° and 90° are shown in Figure 6.5e. The mean thresholds for location discrimination are in good agreement with those for two-point discrimination at the more medial locations (10.4°/9.2° at 0° azimuth; 21.0°/20.2° at 45° azimuth). Note also that these values were in reasonable agreement with Perrott's 'concurrent minimum audible angles' measured using pure tones (Perrott, 1984, see section 3.5.3). However, at the extreme lateral location, the mean threshold obtained in the present speech location discrimination experiment exceeds that obtained for two-point discrimination (51.0°/39.3° at 90° azimuth). As there was such good agreement between the results

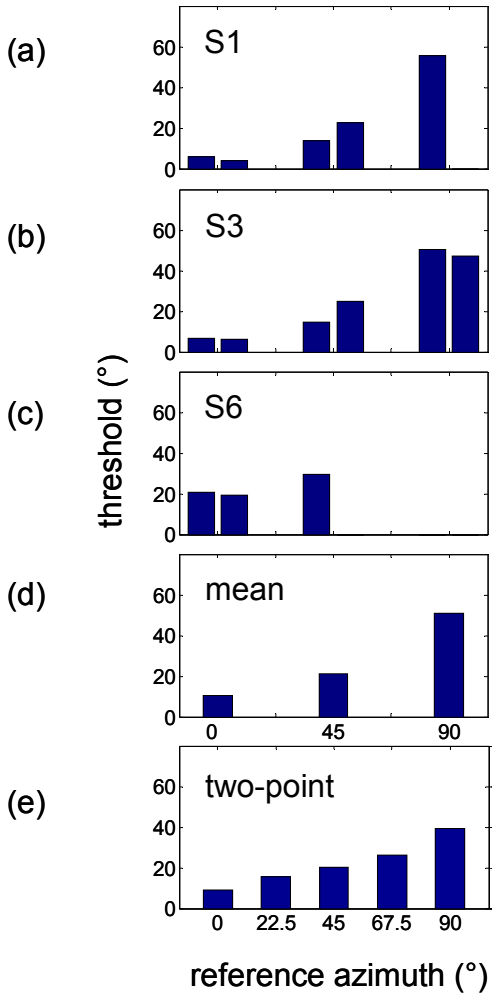



**Figure 6.5** Thresholds for horizontal discrimination of speech stimuli. (a) to (c) show thresholds (estimated from the psychophysical curves in Figure 6.3) for subjects S1, S3 and S6 respectively. For each reference azimuth (0°, 45°, 90°), the left and right bars show thresholds due to leftward and rightward separation. It was not possible to obtain a value in some cases, and thus some bars are missing. The bars in (d) show mean thresholds at each reference azimuth, calculated on the basis of the available individual thresholds. In (e), mean two-point discrimination thresholds (from Chapter 3) for reference azimuths of 0°, 22.5°, 45°, 67.5° and 90° are shown.

for the two experiments at the more medial locations, it is unlikely that this difference is due to stimulus characteristics (i.e. speech versus noise). A more likely explanation for the discrepancy is that for location discrimination around 90° azimuth it is necessary to make accurate front/back judgements (this was discussed above in section 6.2.2) and subjects had trouble with this requirement. In the two-point discrimination experiment there was no such requirement as no location judgements were required.

### Vertical speech discrimination

For vertical separation, it was not useful to calculate thresholds for location discrimination as performance was so poor overall, even with broadband stimuli. The poor ability of subjects to spatially discriminate on the median vertical plane (reference azimuth 0º) is reminiscent of the results for vertical two-point discrimination on this plane. It was concluded in Chapter 3 that spectral cues are not sufficient for distinguishing two similar broadband noises when no binaural differences are present. The results here for concurrent speech stimuli extend this conclusion. Even if the concurrent broadband stimuli are distinct in their spectro-temporal characteristics, and hence distinct from each other, their locations cannot be resolved on the basis of spectral cues alone. Note however that S3 did show some ability to discriminate on the median vertical plane, although rather poorly and only for low locations. It seems that this subject was able to make use of a spectral cue in some cases. Note also that this ability was not affected by low-pass filtering, suggesting that the cues are likely to be in the low-frequency region (recall too the ability of this subject to localise speech was not profoundly disturbed by low-pass filtering, Chapter 5). In the main, however, it seems that the auditory system is unable to extract both directional spectra from the mixed signal it receives at each ear.

Two subjects (S1 and S3) showed some improvement at the more lateral locations. In Chapter 3 it was seen that separating concurrent noises in elevation at these locations produced good detection of the two sources that is likely driven by ITD. However, the ITD cue is ambiguous in that a given change in ITD can indicate either upwards or downwards displacement. Spectral cues must have driven successful discrimination in these cases. Thus it seems that, given a difference in ITD, spectral cues can be useful cues for location discrimination.

### The influence of bandwidth

It was somewhat difficult to observe the impact of low-pass filtering on this task as performance was quite poor overall. Nonetheless, there were observable disruptions and these occurred in cases where spectral cues were vital in maintaining performance. This includes horizontal discrimination at 90° where front-back differentiation was required (e.g. S1, Figure 6.3a, bottom panel) and vertical discrimination at 45° and 90° where up-down differentiation was required (e.g. S3, Figure 6.4b, middle and

bottom panels). In fact the impact of low-pass filtering on performance in these cases supports the notion that spectral cues were driving the responses. As high-frequency spectral cues are required for the unambiguous localisation of a single speech source (Chapter 5), it is not surprising that they are also required for the unambiguous *relative* localisation of two speech sources.

## 6.3 Experiment 2: Speech localisation in the presence of a concurrent speech masker

### 6.3.1 Experimental methods

**Subjects and task**

This experiment followed the same format as the experiments described in Chapter 4 for localisation in the presence of a masker. The same four subjects were used (S1, S3, S5, S6) as in the previous set of experiments. Experiments were carried out in the anechoic chamber and stimuli were presented in virtual auditory space (Chapter 2). The experiment comprised a block of five localisation tests that were completed in succession (this took approximately 1 hour per subject).

The localisation response paradigm was the same as that described in section 2.4.1. Briefly, subjects were positioned in the centre of the chamber, with their heads calibrated to be facing directly ahead (stimulus position 0°, 0°). Stimuli were presented over headphones (stimulus positions described below) and subjects pointed with their noses to their perceived location of the target source. An electromagnetic head-tracker recorded their estimates after a response button was pressed

**Stimuli**

Pairs of words were chosen in a similar fashion to Experiment 1. For each of the five tests, a target word was chosen randomly from the corpus of monosyllabic words described in section 5.3.2. This word was identified to the subject at the beginning of the test, and it served as the target for the 92 trials of the test. The masking word was kept constant throughout the test, and was chosen from the corpus such that it matched the target word as closely as possible in duration (see Figure 6.1 for an illustration). The generation of target/masker stimulus pairs was identical to that

described for Experiment 1. The two words were of approximately equal level, giving an overall signal-to-noise ratio of approximately 0 dB. After summing, stimuli produced a sensation level of approximately 50 dB.

Stimulus configurations were as described in Chapter 4 (section 4.4.2 and Figure 4.1). Stimuli were presented in the frontal hemisphere of space and are described using lateral and polar angle co-ordinates. The masker was presented at one of three masker locations: (-30°, 0°), (0°,0°) and (30°,0°) and the target from one of 19 locations distributed around each masker. Each test consisted of 92 trials, and on each trial a target stimulus was presented either in isolation (35 trials) or in the presence of a masker (57 trials). The five localisation tests resulted in five estimates of the perceived location of all stimuli presented in this experiment. Data were analysed using the approach described in section 4.4.3.

## 6.3.2 Results

### Overall performance

Table 6.1 shows the SCCs obtained for each subject in masker and no-masker trials. S3 and S5 were more accurate localisers, in line with their performance in previous experiments (Chapters 4 and 5). However all subjects localised to a reasonable degree of accuracy, with SCCs in the control condition of 0.66, 0.94, 0.93, and 0.77. In the presence of the masker, it can be seen that all subjects were able to maintain their control level of performance as the SCCs did not alter dramatically (0.64, 0.93, 0.91, and 0.77).

**Table 6.1** Spherical correlation coefficients (SCCs) for each of the four subjects (S1, S3, S5, S6) in control (no masker) trials and test (with masker) trials of Experiment 2. Each SCC was calculated from 175 control trials or 285 test trials. See section 2.4.3 for details of this statistic.

| | S1 | S3 | S5 | S6 |
|---|---|---|---|---|
| cont | 0.66 | 0.94 | 0.93 | 0.77 |
| test | 0.64 | 0.93 | 0.91 | 0.77 |

### Qualitative analysis of responses

Although the masker did not dramatically alter overall performance levels, it was observed to influence estimates of target location. Response patterns can be observed in Figure 6.6. Each row shows results from one of the four subjects. The three spherical plots in a row show data for the three masker locations (left: 30°,0°; middle: 0°,0°; right -30°,0°). The masker location is depicted by a large red dot and the location directly in front of the listener is indicated by the grey circle or 'nose'. Centroids of localisation estimates in the presence of the masker (blue dots) are joined by a blue line to corresponding control centroids (red dots).

For all subjects, it can be seen that the effects induced by the masker are relatively small. Most centroid pairs are located close together in a region that appears to be well-differentiated from other target locations. However, there is evidence of a consistent lateral angle bias akin to that observed for non-speech stimuli in Chapter 4. Although there is quite some scatter in the data, close inspection reveals that the masker tends to 'push' the perceived target lateral angle away from that of the masker. This effect is stronger in some subjects and some locations. The masker also appears to have an effect on perceived target polar angle, although this effect is non-systematic. In two subjects there is very little polar angle disturbance and most of the shifts are directed laterally (e.g. S3 and S6). In the other subjects (S1 and S5) polar angle shifts can be seen between masker and no-masker pairs, and these are occasionally quite large, but no systematic pattern is apparent.

### Quantification of localisation bias

The differences between masker and no-masker centroids were calculated and expressed as lateral and polar angle shifts. These shifts were expressed relative to the masker lateral or polar angle, such that a positive shift indicates a bias away from the masker and a negative shift indicates a bias towards the masker. Data from all masker locations was pooled, but individual subjects were treated separately in order to capture the variability seen in the spherical plots.

Figure 6.7 illustrates the mean lateral and polar angle shifts (and SEMs). The different coloured bars represent the individual subjects. The first thing to note is that all mean lateral angle shifts are positive, indicating a bias away from the masker lateral angle. The asterisks indicate shifts that were significantly different from zero,

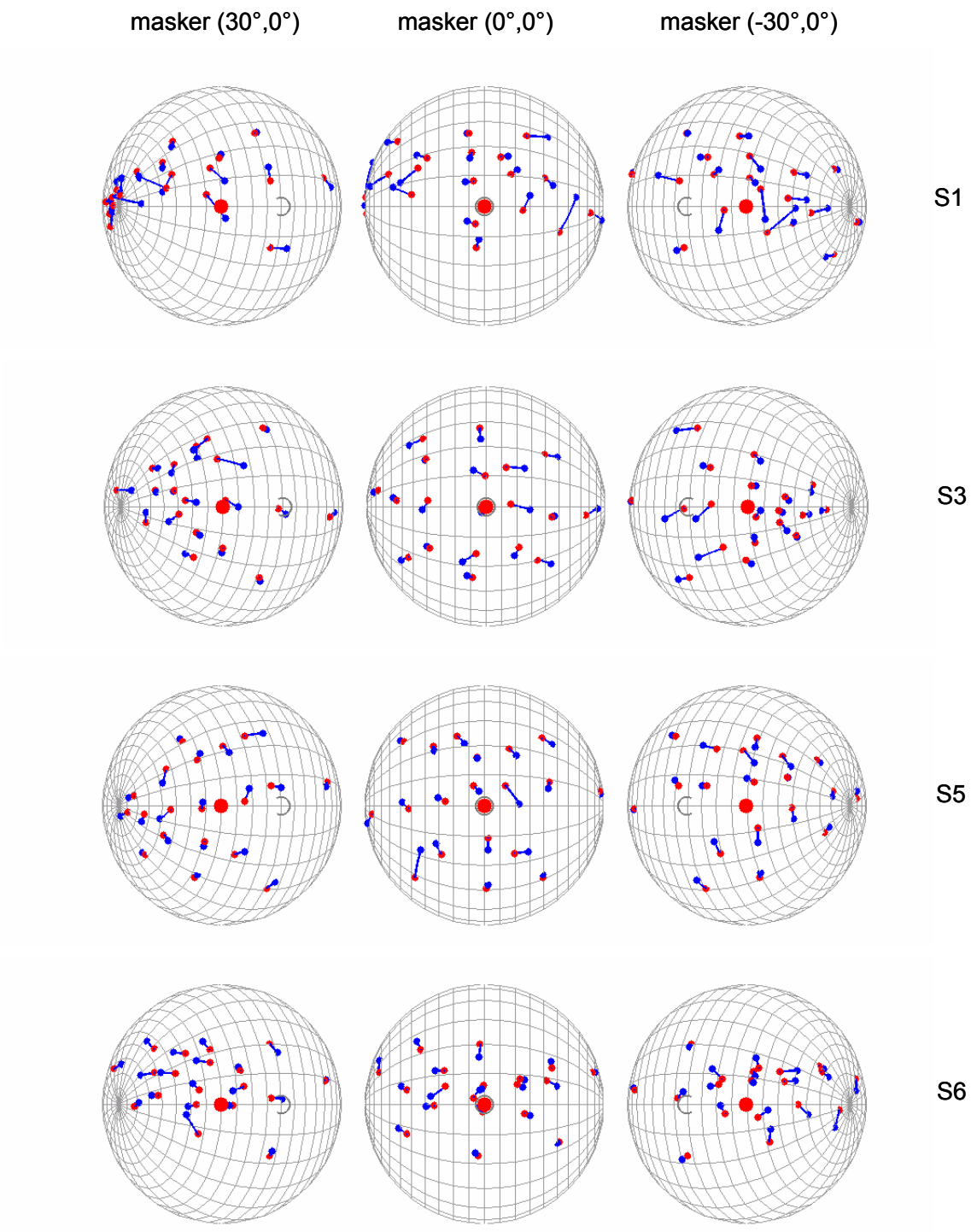



**Figure 6.6** Spherical plots for the concurrent speech stimuli of Experiment 2. The four rows depict results from the four subjects, and the three spherical plots in a row show data for the three masker locations (left: 30°,0°; middle: 0°,0°; right -30°,0°). On each plot the masker location is depicted by a large red dot and the location directly in front of the listener is indicated by the grey circle or 'nose'. Red dots show the centroids of localisation estimates in the control condition. Joined to these by blue lines are blue dots representing the centroids of corresponding estimates in the presence of the masker.

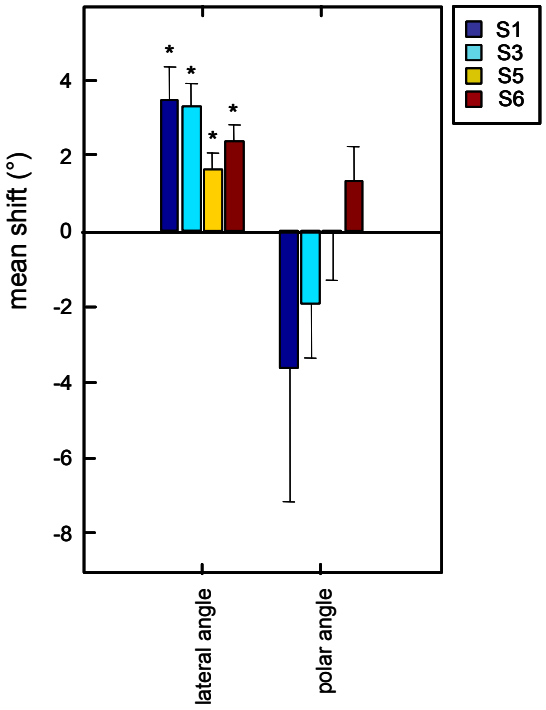



**Figure 6.7** Mean lateral and polar angle shifts (across all locations and maskers) in Experiment 2. Each bar in a group represents a different individual. A positive shift indicates a shift away from the masker lateral (or polar) angle. Error bars show standard error of the mean, and asterisks indicate shifts that were significantly different from zero ($p<0.05$).

which was the case for all subjects. The mean polar angle shifts were both negative and positive, depending on the individual, and the large SEMs indicate the high variability in the data. In no subject was the mean polar angle shift significantly different from zero (two-tailed t test, $p>0.05$).

### 6.3.3 Discussion

It was found in this experiment that while subjects make some errors when localising a single broadband word, their estimate of location is fairly robust in the presence of a simultaneous word.

As speech has such a dynamic spectral structure, it is particularly remarkable that spectral cue (polar angle) localisation was maintained. In Chapter 5 it was seen that speech localisation is extremely vulnerable to the loss of high-frequency information, and yet it appears relatively insensitive to the presence of simultaneous

interference. It is almost certain that there would have been overlap between the competing sources in this important high-frequency region, but somehow subjects were able extract the appropriate spectral cues for the target. This process is likely to be contingent on an accurate segregation of the two sources, and it is well known that speech contains very strong grouping cues (see Darwin and Carlyon, 1995 for a detailed discussion). Certainly subjects reported a clear distinction between the two sources and a clear perception of each word, suggesting good segregation did occur. However it is still remarkable, even assuming the successful segregation of the two words, that the high-frequency spectral cues specifically associated with *direction* can be extracted from such a complex signal. This brings us back to the discussion given in Chapter 1 (section 1.3.3) where it was noted that the spatial auditory system seems to have a solution to the problem of confounded source and directional spectra in natural situations (see also Rakerd *et al.*, 1999).

Although both lateral and polar angle of the target word were well maintained in the presence of the masking word, there were some small shifts induced by the masker (Figures 6.6 and 6.7). Note that the pattern of these effects was very similar to that seen for flat spectrum broadband stimuli examined in Chapter 4. Lateral angle shifts indicated a consistent bias *away* from the lateral angle of the masker, whereas polar angle shifts appeared to be unsystematic. Furthermore the lateral angle bias was relatively constant across location, regardless of the distance between target and masker (see Figure 6.8) and this too has been discussed (see section 4.7.2 and Figure 4.17). This correspondence between speech stimuli and noise stimuli in terms of spatial interactions is good evidence that the effects observed are not stimulus-specific, and are likely to occur for any pair of sounds that overlap largely in frequency and time.

## 6.4 General discussion

A very interesting observation can be made by comparing the results from the two experiments described above. In Experiment 1, it was found that subjects had trouble discriminating the locations of concurrent words separated along the median vertical plane (MVP). This is presumably related to the fact that there are no binaural differences between the stimuli, i.e., the words are located on the same cone of

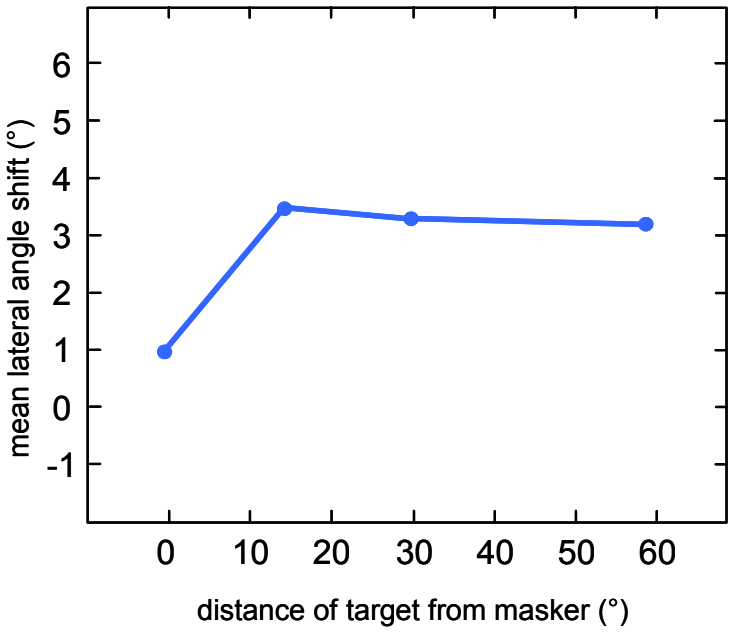



**Figure 6.8** Effect of distance from the masker on lateral angle localisation bias with concurrent speech stimuli. Responses have been pooled across subjects and locations and the mean lateral angle shift is plotted against the distance (in lateral angle) of the target from the masker.

confusion. In Experiment 2, there were many instances in which both words were presented on the MVP, or indeed shared the same cone of confusion at a non-zero lateral angle. Subjects were asked to localise just a target word in this experiment, and subjects performed well. Taken together, these results are surprising: on the MVP subjects could localise the *target* source in a concurrent pair, but could not localise *both* sources in order to judge relative location.

This discrepancy is interesting and deserves some consideration. When a target source was designated and subjects were able to ignore the masker, localisation was quite accurate. There is no reason why the masker word could not have been designated as the target, and localised just as well. However, during the stimulus presentation, it seems that only one stimulus could be accurately localised. Importantly this discrepancy was not apparent for configurations in which binaural differences were present. For example, in Experiment 1 subjects showed a good ability to discriminate horizontal location, suggesting that both sources were localised 'binaurally' with good precision.

A possible explanation for the loss of ability in the MVP discrimination experiment is that subjects did not have enough *time* during the stimulus presentation

to accurately gauge the location of both stimuli. It may be that localisation (via spectral cues) of a speech sound takes too much time to be performed twice within the duration of the stimulus. Indeed spectral cues are processed more slowly than binaural cues according to recent auditory evoked magnetic field work (Fujiki *et al.*, 2002). Furthermore, this notion has been validated behaviourally. Using broadband noise, Hofman and Van Opstal (1998) showed that optimal localisation in elevation requires a longer stimulus duration (80 ms) than optimal localisation in azimuth (3 ms). Furthermore, Strybel and Fujimoto (2000) measured minimum audible angles (MAAs, see section 1.4.2) for noise bursts of different durations in horizontal and vertical planes at (0°,0°). They found that horizontal and vertical MAAs were similar (about 4°) until stimulus duration was reduced to 10 ms, when only vertical performance dropped (to about 12°). A problem with this explanation is that these durations are an order of magnitude smaller than the duration of stimuli in the current experiments (about 700 ms on average). However, the spectral analysis that has to take place to localise two spoken words on the MVP is more involved than that required in these previous studies. In the current experiment, the analysis involved in (a) identifying the target, (b) extracting spectral cues corresponding to the target, (c) identifying the masker, and (d) extracting spectral cues corresponding to the masker. Given the complexity of the task, it follows that while the horizontal and vertical differential remains, a longer absolute duration is required.

An alternative interpretation is that location processing was not the limiting factor, but that subjects were unable to effectively *attend* to each stimulus in the MVP discrimination task. There are two possible realisations of this hypothesis. Firstly subjects may adopt a 'divided attention' strategy, and attempt to monitor the two locations simultaneously. In this case, the failure on the MVP may reflect an attentional system that can only be divided along the primary binaural dimension. On the other hand, it may be that subjects adopt a 'selective attention' strategy, where one object is attended and then attention is 'switched' to the other. In this case it may again be that duration is the limiting factor; it may take a prolonged time to effectively attend and switch attention along the MVP given the complexity of the spectral analysis as described above.

## 6.5 Conclusions

Two experiments were performed to examine spatial performance with concurrent speech sources. In Experiment 1, using a location-discrimination approach, it was found that horizontal resolution for speech worsened with increasing laterality. In the vertical dimension, discrimination was poor, particularly when no binaural differences were present. In Experiment 2, listeners were required to judge the location of a designated target in concurrent speech pairs. Localisation was reasonably unaffected by the competing source, even when no binaural differences were present. This was intriguing since subjects could not localise *both* sources in such a pair for the purpose of location discrimination in Experiment 1. The apparent discrepancy may be related to attentional capabilities.

# Chapter 7: Spatial factors aiding speech segregation

## 7.1 Introduction

In Chapters 3 and 4, the effectiveness of the different spatial cues for resolving the locations of concurrent broadband sound sources was examined. In Chapter 6 this investigation was extended to mono-syllabic spoken words in order to measure spatial performance with more acoustically complex and ecologically important stimuli. In this final experimental chapter, the role of spatial cues in the segregation of running sentence speech is considered.

It is generally believed that two types of masking occur in multiple source scenarios. The first is the classic form, known as 'energetic masking', where stimuli overlap in frequency and time. In this case the signals interfere acoustically with each other, and this can render a target inaudible. The second is known as 'informational masking' and refers to a situation where competing signals do not necessarily overlap in frequency or time, and may be clearly audible. However if the signals are similar in some other way, uncertainty is increased and confusions can occur between the competing streams.

Energetic masking has been studied intensively, with tasks ranging from detecting tones in noise to understanding speech in the presence of noise or competing speech. It has been shown that binaural factors play an important role in such situations. If a target is undetectable or unintelligible when co-located with a concurrent masker, then horizontal separation of the pair can enable detection or intelligibility. This phenomenon is known as 'binaural release from masking' and is thought to have two primary components (Dirks and Wilson, 1969; Zurek, 1990). Firstly, acoustic shadowing by the head influences the levels of the target and the masker at the two ears. Specifically, for sound sources located off the midline, the 'head-shadow' results in an improved signal-to-noise ratio at *one ear*. However, binaural listening is better than can be predicted on the basis of monaural listening at

the better ear. This added advantage comes from 'binaural interaction', whereby differences in ITD and/or ILD lead to increased target detectability (Bronkhorst and Plomp, 1988).

Informational masking has received increasing amounts of attention in recent years. It has been commonly examined via the detection of tone patterns in the presence of competing tone patterns that do not overlap in frequency (e.g. Kidd *et al.*, 1998). However more recently informational masking has been reported to play a role in multiple talker situations, when interfering words can disrupt a listener's ability to follow a target talker. It appears that binaural separation can also reduce this form of masking. This may be due to a reduction in confusion or uncertainty provided by the perceived differences in location of the competing sources (e.g. see Freyman *et al.*, 1999; Freyman *et al.*, 2001; Brungart *et al.*, 2001).

The situation of interest in the present experiment was when there are several talkers delivering different streams of speech at the same time. Both forms of masking as described above may be present in this situation. The experiment was aimed at further investigating the influence of spatial separation on the segregation of competing speech sources. Segregation is defined here as the assignment of the correct word content to the appropriate talkers, and is measured by one's ability to correctly recall the words spoken by a target talker.

Clearly a binaural advantage was expected, as many authors have reported such an effect. However it was of interest to determine whether spatial separation that does *not* involve binaural separation could also afford an advantage. The role of spectral cues in spatial hearing with concurrent sources was addressed in previous chapters. Spectral cues were shown to be of use in some situations (see Chapter 4; Chapter 6 Experiment 2) but not others (see Chapter 3; Chapter 6 Experiment 1). In the present experiment, competing speech sources were separated along the median vertical plane to examine the role (if any) of spectral cues in selective listening. As in previous experiments (Chapters 5 and 6), broadband and low-pass filtered versions of the speech stimuli were used to specifically examine the role of high-frequency spectral cues.

Another aspect investigated in this experiment was that of voice characteristics. In the previous chapter, the competing speech signals were always spoken by the same voice. This increased the overlap in frequency and increased 'energetic' effects. However difference in voice pitch is a clear cue for segregation in natural

environments (e.g. Brokx and Nooteboom, 1982). Thus in this experiment, the opportunity was taken to examine possible interactions between voice and spatial factors in the segregation of running speech.

## 7.2 Experimental methods

### 7.2.1 Subjects and task

Six subjects (S1, S3, S4, S8, S9 and S10) participated in the experiments. S8, S9 and S10 were relatively inexperienced subjects, having participated only in the localisation training and virtual auditory space validation regime described in Chapter 2. The other four subjects had previous experience in auditory psychophysical experiments. The experiment was carried out in the soundproof room and stimuli were presented in virtual auditory space (see Chapter 2).

On each trial the subject was presented with three competing sentences, and was asked to identify two keywords spoken by a target talker whilst ignoring words spoken concurrently by the other two talkers (details in next section). The subjects were familiarised with the speech materials and the task by completing a practice test (of 36 trials, see section 7.2.5) prior to testing.

### 7.2.2 Speech materials

The speech materials consisted of spoken sentences that were taken from the publicly available Coordinate Response Measure (CRM) corpus (Bolia *et al.*, 2000). These sentences all contain seven words, three of which are variable keywords. The form of the sentences is "Ready *call-sign* go to *colour number* now", where the italicised words indicate keywords. In the corpus there are eight possible call-signs ("arrow", "baron", "charlie", "eagle", "hopper", "laker", "ringo", "tiger"), four possible colours ("blue", "green", "red", "white"), and eight possible numbers (1-8). All combinations of these words gives 256 phrases, which are each spoken by eight talkers (four male, four female), giving a total of 2048 available sentences. The sentences are time-aligned such that the word "ready" always starts at the same time, but some variations in overall rhythm occur between different sentences. As the signals in the corpus were

recorded at a sampling rate of 40 kHz, they were up-sampled to 80 kHz in order to be compatible with the DTFs used in the virtualisation.

Three different sentences were presented concurrently on every trial. One of these was designated the target, and always contained the call-sign "baron". The other two sentences were chosen to have different call signs, different colours and different numbers from the target and each other. The three sentences were spatialised and summed without intensity scaling, such that all three talkers were at an approximately equal conversational level.

### 7.2.3 Stimulus configurations

Stimulus locations are described in this chapter using the lateral/polar angle co-ordinate system (see section 2.2). The target was always presented from the location directly in front of the subject (0°,0°). The two maskers were both presented from a single location. Nine masker locations were employed in the right hemisphere of space. One of these was (0°,0°), meaning that for these trials all three talkers were co-located at this location. Four masker locations were chosen on the audiovisual horizon and differed from the target location in lateral angle only (Figure 7.1a). Their lateral/polar co-ordinates were: (10°,0°), (45°,0°), (90°,0°), (45°,180°). Three masker locations were chosen on the median vertical plane (MVP) and differed from the target in polar angle only (Figure 7.1b). Their lateral/polar co-ordinates were: (0°,10°), (0°,45°), (0°,180°). The final masker location differed from the target location in both lateral and polar angle (45°,45°).

### 7.2.4 Stimulus conditions

For all of the stimulus configurations described above, two stimulus issues were considered. First, the effect of bandwidth was investigated by comparing performance with broadband stimuli ('broadband') and stimuli low-pass filtered at 8 kHz ('low-pass'). The publicly available corpus is band-limited to 8 kHz and was used for the latter condition. For the broadband condition, the original recordings were obtained from the creators and low-pass filtered at 16 kHz. Second, the effect of voice characteristics was examined by comparing performance under 'same talker' and

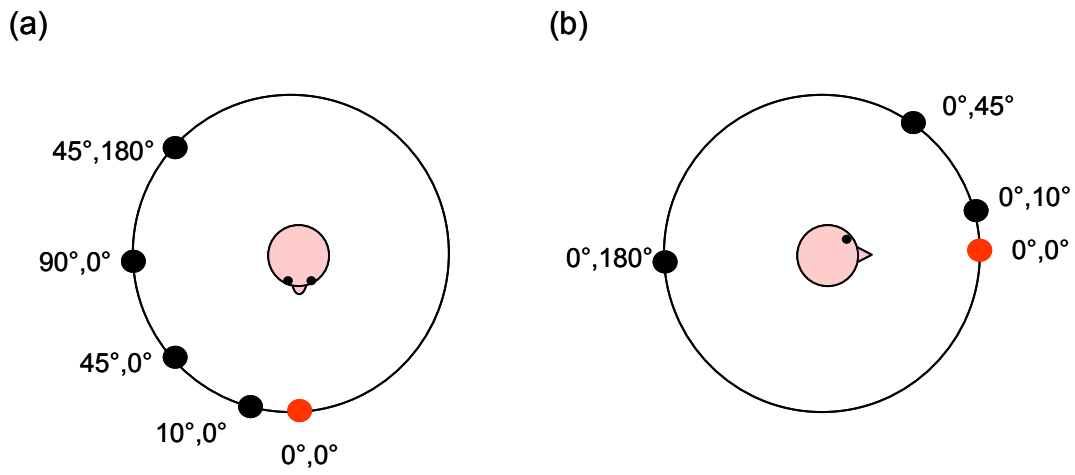



**Figure 7.1** Stimulus configurations, described using the lateral/polar co-ordinate system. The target (red dot) was always located at (0°,0°). Masker locations (brown dots) were separated in lateral angle (a) or polar angle (b) or both (not shown).

'different talker' conditions. In the 'same talker' condition, all three sentences were spoken by the same voice although the voice could vary randomly from trial to trial. In the 'different talker' condition, the target voice was chosen randomly on each trial and the two masker voices were chosen pseudo-randomly to be different from the target and each other.

### 7.2.5 Testing procedure

The four resulting conditions (each combination of two bandwidth conditions and two talker conditions) were interleaved randomly throughout testing. In a single test, each of the 9 stimulus configurations was tested under each condition once, giving a total of 36 trials. 20 of these tests were completed by each subject, giving a total of 20 responses for every stimulus condition/configuration. For each of these replicates, the talkers were drawn randomly from the corpus.

At the beginning of each testing session, the subject was instructed to attend to the talker uttering the keyword "baron", who would be located at the frontal location. On each trial, they were presented with a stimulus and their task was to track the target talker and report back the colour and number uttered by him/her. They responded by entering the first letter of the colour and then the digit representing the number into the laptop computer provided. For example, to respond "red four" the

subject would enter “r4” into the keyboard. An interactive MATLAB script prompted the subject and fed responses via the serial port to the PC controlling the experiment outside the chamber.

### 7.2.6 Data analysis

Responses to the 20 repetitions for each stimulus were collected and analysed to find the percentage correct in each case. A response was deemed correct only if both the colour and number were reported correctly. For each stimulus configuration in which the maskers were separated from the target, performance was compared to the case in which the maskers and target were co-located. This gave some indication of the benefit provided by spatial separation in the chosen dimensions.

When rating performance on a percentage correct basis, it is important to define what constitutes a ‘chance’ level of performance. Unfortunately this is not straightforward for the stimuli and task used in this experiment. The subject had to correctly identify the target colour and number from those possible in the corpus. In the simplest view, they had a 1/4 chance of guessing the colour and a 1/8 chance of guessing the number, giving a 1/32 chance of guessing the pair (about 3%). However, if one or more of the sentences were audible, then the situation changes. For example, it may be that the target was inaudible but both maskers were audible. In this case the subject would exclude the colours and numbers spoken by the maskers before guessing. This would leave two possible colours and six possible numbers, giving a chance performance level of 1/12 (about 8%). As this particular task was designed to measure segregation and not intelligibility, all talkers were presented at equal and favourable levels. Thus it is in fact likely that all three sentences were audible on a given trial, and that performance disruption occurred due to confusion between the different streams. Assuming this is the case, the subject would have heard three colours and three numbers upon which to base their response, raising the chance performance level to 1/9 (about 11%).

# 7.3 Results

## 7.3.1 Same talker condition

Results for the ‘same talker’ trials are shown in Figure 7.2. The first six panels show data for the six different subjects and the final panel shows the mean data. The 9 groups of bars in each panel represent the different testing configurations as labelled. For each configuration, the blue bar shows percentage correct in the broadband condition and the red bar shows percentage correct in the low-pass condition.

Performance in the broadband condition will be discussed first. When the maskers were co-located with the target (0°,0°) it can be seen that overall scores were quite low. Subjects achieved 21% accuracy on average with no spatial separation of the talkers. When the maskers were displaced laterally by 10° (10°,0°) performance improved substantially with a mean of 75%. When the maskers were displaced even more laterally to positions of (45°,0°), (90°,0°), and (45°,180°) performance levels were close to the upper limit (95%, 93%, 94%). Clearly the task was not difficult enough to allow differentiation between performance in these favourable configurations. When a 45° polar angle separation was added to the 45° lateral angle separation (45°,45°), performance was actually at the ceiling level (100%). For the configurations in which maskers were separated along the MVP, there were modest improvements for the frontal positions (0°,10°), (0°,45°) with mean scores of 31% and 38%. A larger improvement was found for the rear position (0°,180°) with a mean score of 69%.

An examination of the red bars reveals that there were some effects of low-pass filtering in some subjects, particularly in the co-located configuration and the MVP configurations. However these effects appear to be non-systematic.

Note that several performance levels in this condition are close to the possible chance level of 11% (see section 7.2.6) and thus may represent instances where the subject simply guessed which keywords were spoken by the target talker.

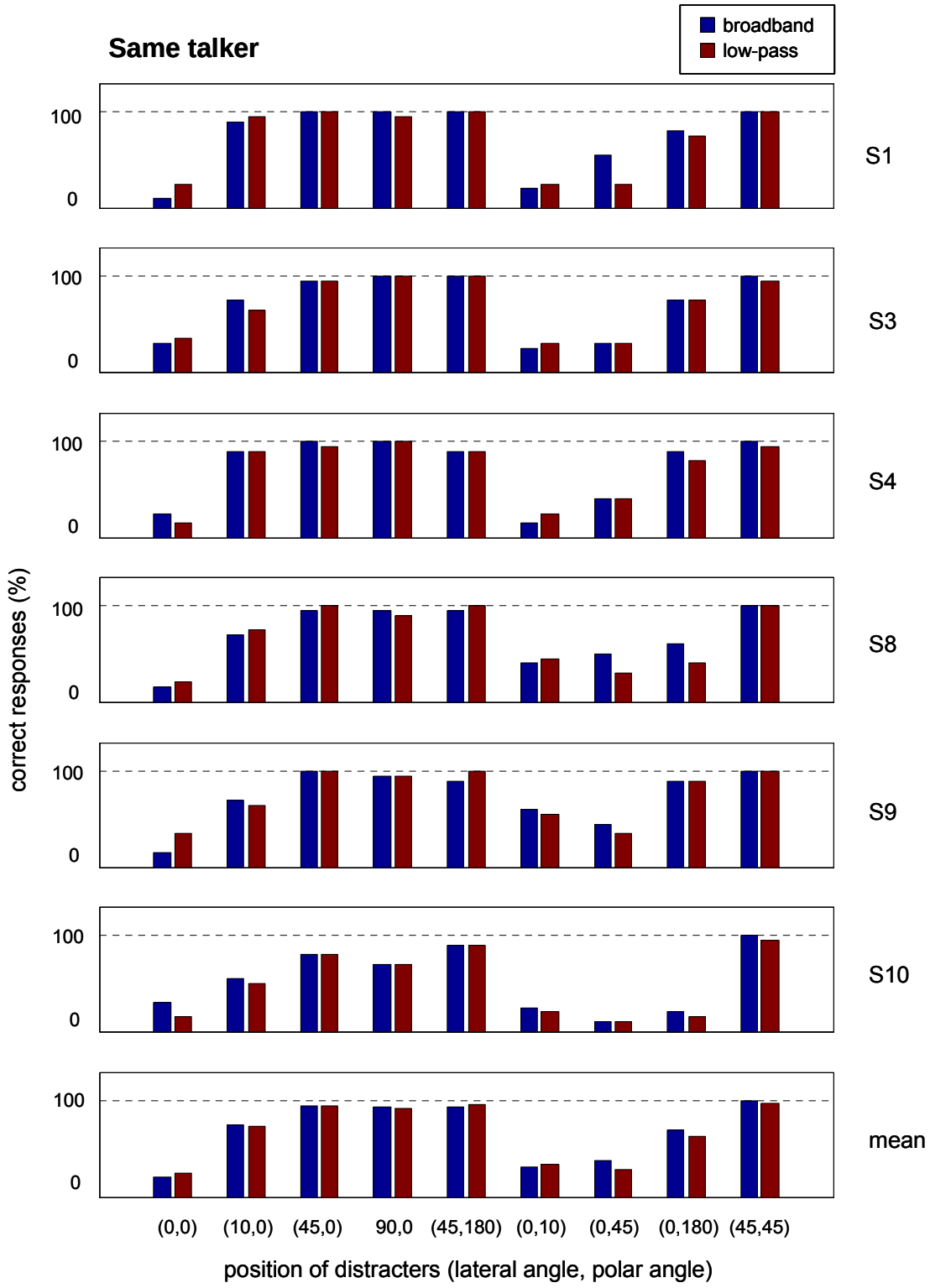



**Figure 7.2** Percentage of correct responses in the same talker condition. The seven rows show data for the six subjects and the mean as labelled. In each panel, results are shown for broadband stimuli (blue bars) and low-pass filtered stimuli (red bars) in the 9 stimulus configurations. Perfect response rate (100%) is indicated by the dotted line.

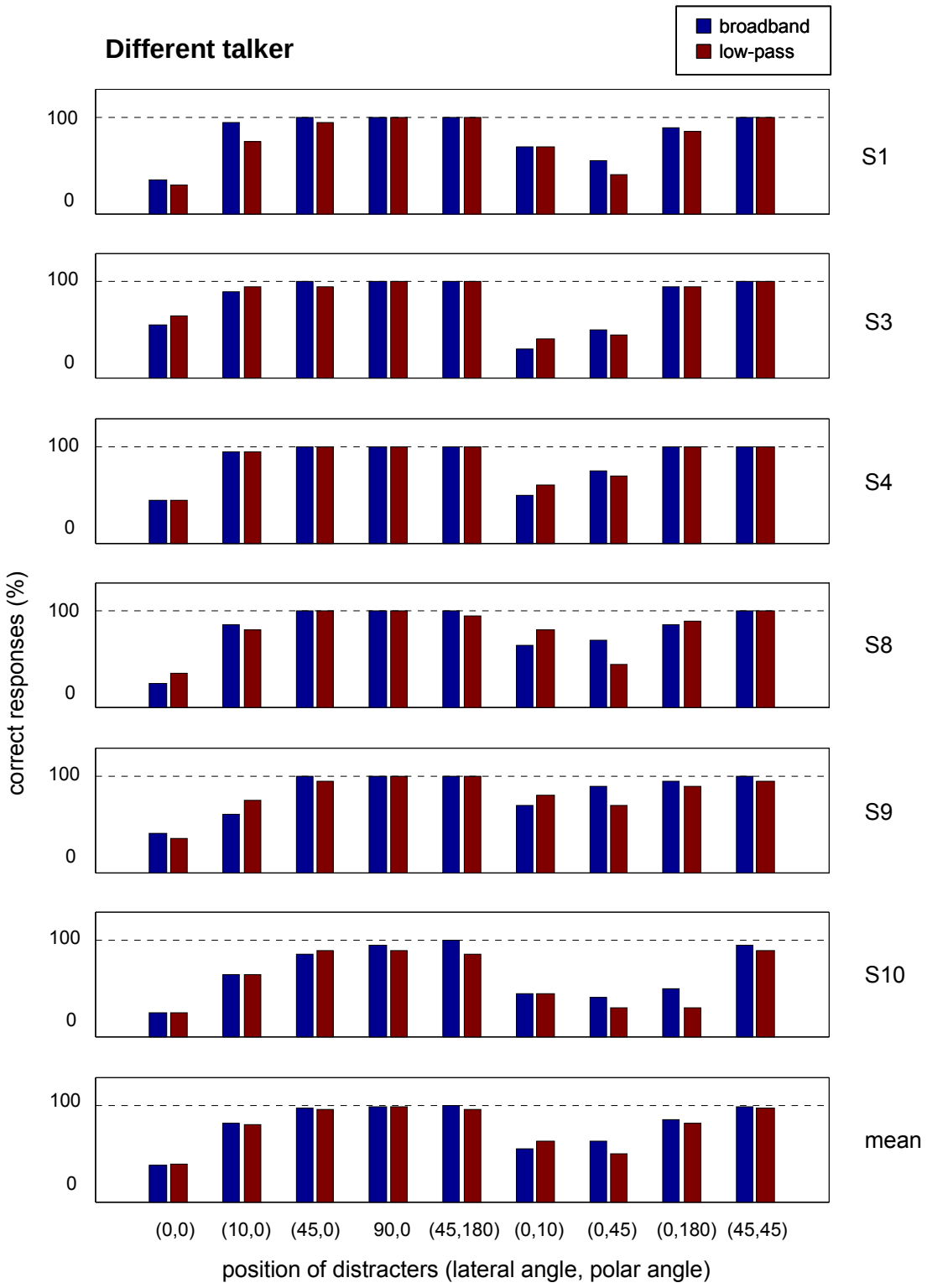



**Figure 7.3** Percentage of correct responses in the different talker condition. All other details as for Figure 7.2.

### 7.3.2 Different talker condition

Results for the ‘different talker’ trials are shown in Figure 7.3, and all details are as for Figure 7.2. Trends in the data are very similar to those in the ‘same talker’ condition, although overall performance is better. Looking again at just the broadband results, it can be seen that when talkers were co-located the mean score was 38% (compared to 21% in ‘same talker’ condition). For maskers displaced laterally by 10°, performance increased to 82% (compared to 75% in ‘same talker’ condition). Again performance saturated for maskers further from the midline with scores of 98%, 99%, 100%, and 99% for masker locations of (45°,0°), (90°,0°), (45°,180°), and (45°,45°).

For the configurations in which maskers were separated along the MVP, improvements were seen that appear to be greater than in the ‘same talker’ condition. Mean scores rose from 38% (co-located) to 55%, 63% and 86% for masker polar angles of 10°, 45° and 180°.

An examination of the red bars reveals that again there were some effects of low-pass filtering, but these were effects were minor. The mean data show a small improvement in performance for the (0°,10°) configuration and a drop in performance for the (0°,45°) configuration using low-pass filtered speech (see next section, and section 7.4.3 for some discussion).

### 7.3.3 Median vertical plane configurations

As there were interesting effects in the MVP configurations, these data are represented for clearer inspection in Figure 7.4.

The top panel shows ‘same talker’ data for broadband stimuli. For each subject the four bars allow a comparison of performance in the four MVP configurations. Although all subjects do not show the same effects for separation of maskers along the MVP, the mean data indicate an overall advantage that increases with increasing separation. Using paired t tests, the advantage was found to be significant at polar angles of 10° [$t(5) = 2.5994$, $p = 0.02$], 45° [$t(5) = 3.7131$, $p = 0.007$], and 180° [$t(5) = 5.2702$, $p = 0.002$]. In the second panel, ‘same talker’ data are shown for low-pass filtered stimuli. The advantage of separation along the MVP

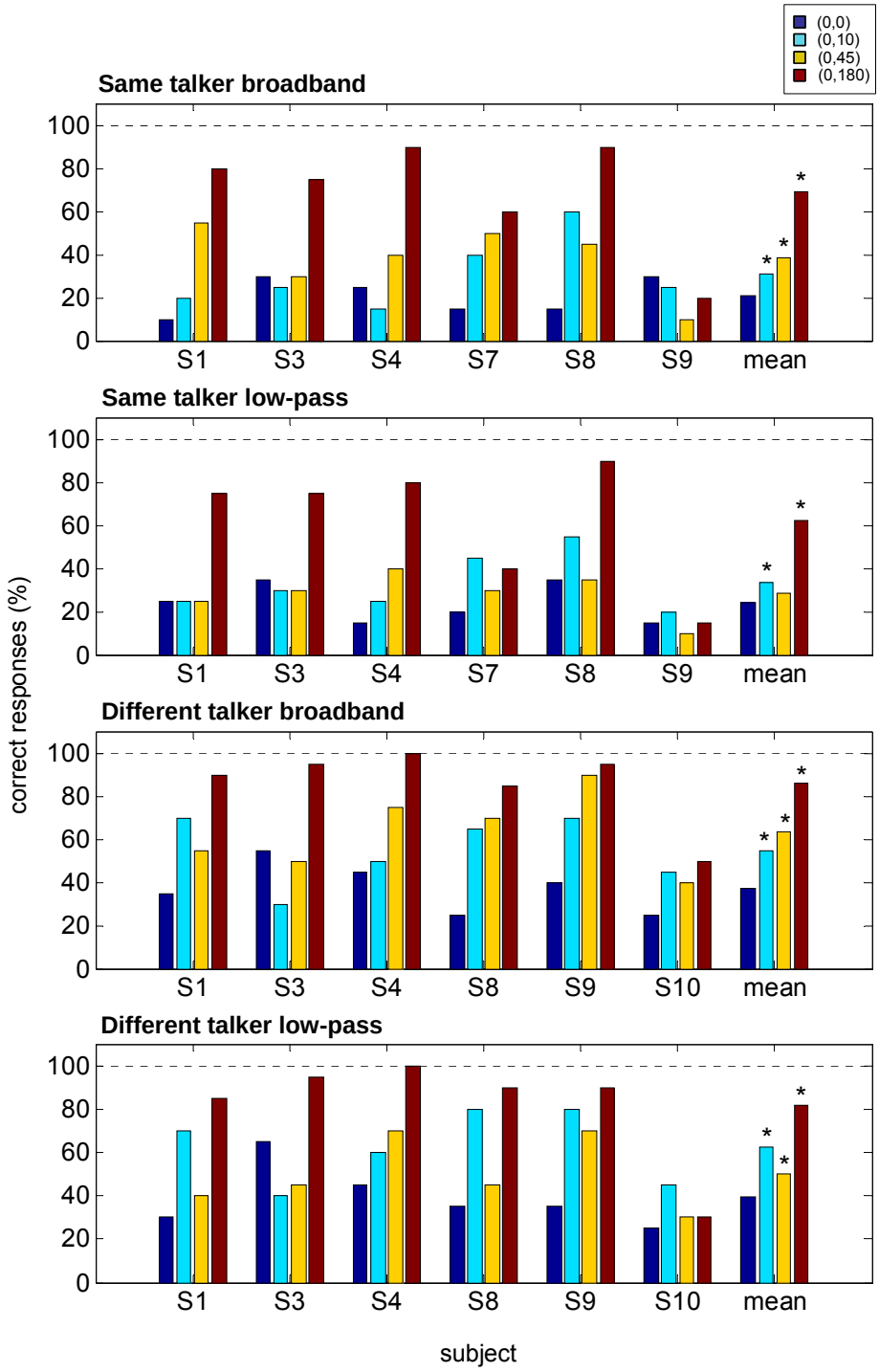



**Figure 7.4** Percentage of correct responses in the median vertical plane (MVP) configurations only (data extracted from Figures 7.2 and 7.3). The four panels show data for the same/different talker condition and for broadband/low-pass filtered stimuli as labelled. Each subject (and the mean data) are represented by a cluster of bars. The dark blue bars depict response rates for the control configuration (masker elevation 0°). The other bars depict response rates for masker elevations of 10° (light blue bars), 45° (yellow bars), and 180° (red bars). Perfect response rate (100%) is indicated by the dotted line. For the mean data, asterisks indicate test configurations in which performance differed significantly ($p<0.05$) from the control (dark blue bars).

persisted for masker polar angles of 10° [$t_5 = 2.7346$, $p = 0.02$] and 180° [$t_5 = 3.8766$, $p = 0.006$], but was lost at 45° [$t_5 = 1.964$, $p = 0.05$].

In the third and fourth panels, 'different talker' data are shown for broadband and low-pass filtered stimuli respectively. For broadband stimuli, all subjects except for S3 show improved performance for all separations. S3 only shows an improvement for maskers located at (0°,180°). The mean data indicate an overall advantage that increases with increasing spatial separation. This advantage was found to be significant using paired t tests at polar angles of 10° [$t(5) = 5.0964$, $p = 0.002$], 45° [$t(5) = 3.8414$, $p = 0.006$], and 180° [$t(5) = 8.9073$, $p = 0.0001$]. For low-pass filtered stimuli, there was again a performance improvement with MVP separation for all subjects except for S3. The advantage was significant at 10° [$t(5) = 5.8358$, $p = 0.001$], 45° [$t(5) = 3.7963$, $p = 0.006$], and 180° [$t(5) = 4.9771$, $p = 0.002$]. Notice however that the effect was reduced at 45° in this case.

## 7.4 Discussion

### 7.4.1 Performance with co-located stimuli

In the configuration in which all three talkers were located at (0°,0°), it was found that subjects reported the keywords correctly in 21% of 'same talker' trials and 38% of 'different talker' trials. These values are both higher than might be expected on the basis of random guessing (see section 7.2.6). Clearly there are non-spatial cues that enable some selective listening to occur. One factor is the signal-to-noise ratio; the target was clearly audible in the presence of the equal-level maskers (SNR approximately -3 dB). Furthermore, voice cues and prosodic features of different speech samples are known to provide strong cues for segregation (Brokx and Nooteboom, 1982; Bregman, 1990; Darwin and Hukin, 2000). The fact that subjects did more poorly in the 'same talker' trials suggests that voice differences offer a large advantage. In particular, pitch information should occur well below 8 kHz, and indeed low-pass filtering at 8 kHz did not greatly disturb performance. It is interesting to point out however that due to the 300 Hz lower cut-off of the DTFs used to spatialise these stimuli, some of the very low-frequency voice information would have been

removed (including the fundamental frequencies). Thus the level of performance observed may in fact be an underestimate of human capabilities in natural settings.

### 7.4.2 Advantage of binaural separation

Although the results for horizontal separation were tempered by a ceiling effect in this experiment, it was clear that binaural separation was advantageous for segregation of a target talker from two competing talkers. It was found that a separation of only 10° produced a marked improvement in performance. Some of the earliest work in this area reported improvements in speech intelligibility at around the same separation (e.g. Spieth *et al.*, 1954). Recall also that the mean threshold for horizontal two-point discrimination at (0°,0°) was found to be 9.2° (section 3.5.3 and Figure 3.5a), and the mean threshold for left/right discrimination with concurrent words was found to be 10.4° at this location (section 6.2.3 and Figure 6.5d).

### 7.4.3 Advantage of median plane separation

The question of most interest in this experiment was whether or not spectral localisation cues could influence the segregation of competing streams of speech. It was found that overall there was some advantage afforded by separating the two maskers from the target along the MVP. When maskers were located at 180°, the improvement in segregation was large for all subjects in all conditions. Results for the frontal MVP masker locations (10° and 45°) were more varied. When all talkers had the same voice an improvement was only seen for some subjects, and when all talkers had different voices an improvement was seen in all but one subject. Overall these improvements were much more modest than those produced with binaural separation.

Only two other studies were found in which the impact of MVP separation on concurrent speech segregation was examined (Worley and Darwin, 2002; McAnally *et al.*, 2003). In both of these studies the testing paradigm was similar but only one masker was used. Accordingly, overall performance was better than in the present study. McAnally and colleagues examined three MVP locations: -50°, 0°, and 50° polar angle. They found a significant improvement when two same-sex talkers were separated rather than co-located (from 64% to 72% correct) and furthermore showed

that this effect was not due to small binaural differences that may occur along the MVP if symmetry is not perfect. In the study of Worley and Darwin, a separation of 31° along the MVP produced almost perfect performance, regardless of the pitch of the talkers. When separation was reduced to 19°, performance worsened, and fell to near chance for talkers with the same fundamental frequency. It seems that listeners in the present study performed better than Worley and Darwin's subjects for small separations. Here a significant benefit of 10° separation was observed in the 'same talker' condition, and this was in the presence of two maskers and not just one.

There are several possible interpretations of the advantage provided by separation on the MVP. The first possibility is that small binaural differences between locations on the MVP may have provided a cue, and this is certainly likely for S4 (recall Figure 3.6). However the findings of McAnally *et al.* (2003) show that these kinds of effects occur even upon removal of such binaural differences. Another possibility is that there may be a small acoustic advantage afforded by different configurations. It is possible that peaks and notches in the spectra imposed on the target and maskers may improve the target-to-masker ratio in certain frequency ranges, effectively increasing the relative salience of the target. Saberi and colleagues (Saberi *et al.*, 1991a) saw a masking release of 8 dB for a separation of 60° along the MVP, indicating that spectral differences can be utilised for target detection. Furthermore, Gilkey and Good (1995) reported that this unmasking only occurs for high-frequency regions of stimuli, which is where the most spectral variation would be expected. Certainly placing maskers at 180° (behind the head) produces a large attenuation in the high-frequency region, and this is where a large improvement was seen in the present study. If high-frequency differences are to explain the present findings however, it is somewhat surprising that low-pass filtering stimuli did not have a dramatic impact on performance. It may be that the benefit derives primarily from mid-frequency regions (below 8 kHz). Although it did appear that the 45° masker location advantage was disrupted in the low-pass condition, and thus for this configuration it may be that a high-frequency feature is important.

In order to examine these possibilities directly, directional transfer functions (DTFs) for the MVP plane positions of interest here were plotted for inspection. In Figure 7.5, the six rows represent DTFs for the six subjects. The three panels in a row show different location pairs to enable a comparison of the overall relative gains. In

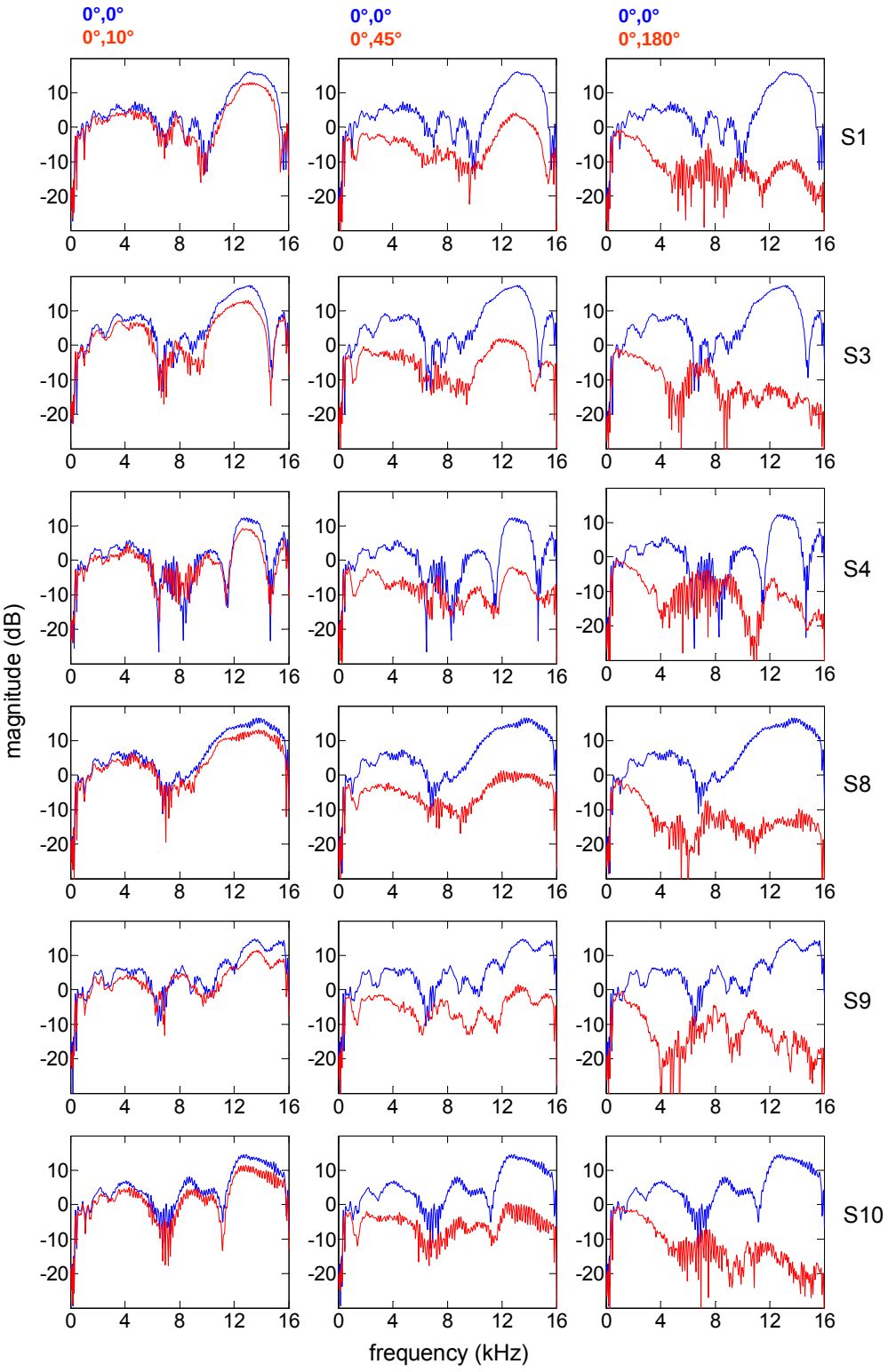



**Figure 7.5** Comparison of left ear directional transfer functions (DTFs) for MVP stimulus location pairs. DTFs for the six subjects are shown in the six rows. Each panel shows the DTF for (0°,0°) in blue. The red curves in the three columns represent DTFs for the three MVP test locations: left (0°,10°), middle (0°,45°), and right (0°,180°).

the left hand panels, DTFs for (0°,0°) and (0°,10°) are shown in blue and red respectively. It can be seen that some decrease in gain occurs for most frequencies for the 10° increase in elevation. This suggests that moving the maskers to (0°,10°) may improve the target-to-masker ratio by a small amount (a few dB). In the middle panels, DTFs for (0°,0°) and (0°,45°) are shown in blue and red respectively. It can be seen that a substantial decrease in gain occurs in the low (<8 kHz) region and an even greater decrease in gain occurs in the high (>8 kHz) region with the 45° increase in elevation. This may explain the benefit obtained by moving the maskers to (0°,45°), and may also explain why low-pass filtering reduces this improvement. In the right hand panels, DTFs for (0°,0°) and (0°,180°) are shown in blue and red respectively. Here the large relative attenuation for the rear position is clear, especially in the high-frequency region. In total, the results are reasonably consistent with the idea that an acoustic advantage when separating source on the MVP can improve segregation. As a final note, individual differences in spectra and hence the acoustical effect might explain individual differences in performance in MVP configurations. However, these individual effects are not apparent in an inspection of the DTFs. For instance, it is not clear why all subjects except S3 would derive a benefit from separation on the MVP. The DTFs for this subject do not appear to differ in any special way from the other subjects. It may be that moment-to-moment variations in level, not captured by these figures, are the key factor. It may also be that subjects weight different aspects of the signal in an individualised way.

An alternative explanation for the influence of spectral cues on speech segregation is that the formation of distinct spatial objects improves the ease with which listeners can selectively attend to the target. In other words, it may be the *perceived* spatial separation of the target and masker that reduces confusion between the two. The work of Freyman and colleagues has most clearly illuminated this idea (Freyman *et al.*, 1999; Freyman *et al.*, 2001). In this work, the precedence effect was used to create the *illusion* of spatial separation of competing speech sources, whilst minimising acoustic advantages due of head-shadow and binaural interaction. It was found that this perceived separation produced an advantage for speech intelligibility. Using a different approach, Driver (1996) used the ventriloquist effect to manipulate perceived target location without changing the acoustic signal. Competing signals were played simultaneously from a single loudspeaker, and a video screen with lip

movements appropriate to the target was located either at the loudspeaker or at a displaced dummy loudspeaker. When displaced, subjects were better able to attend the target, presumably because the ventriloquist effect 'dragged' the perceived location of the target away from the distracter. 'Informational unmasking' has also been described in terms of a reduction in stimulus uncertainty (see Durlach *et al.*, 2003 for discussion). This idea may explain why the subjects in the present experiment performed better than those of Worley and Darwin (2002) for closely spaced MVP locations. In the latter study, the target location was randomised between the upper and lower location on each trial, whereas listeners in the present study were encouraged to attend to the frontal location, eliminating one source of uncertainty.

### 7.4.4 Types of masking and voice/space interactions

It is likely that the two forms of masking described in section 7.1 occurred in the stimulus arrangement used in this study. Energetic masking is inevitable as speech is relatively broadband (see Chapter 5) and spectral overlap will occur with three competing talkers. Informational masking has been explicitly identified in experiments using the CRM paradigm (Brungart *et al.*, 2001); incorrect responses are most commonly due to the reporting of masker keywords instead of target keywords. Importantly, the two mechanisms described in the previous section by which spectral cues may enhance segregation are intimately related to these two forms of masking, and thus *both* of the identified mechanisms may well have played a role.

The idea that perceived separation plays an important role in speech segregation has received growing support and is thought to exert its influence primarily by reducing *informational* masking (see Brungart and Simpson, 2002; Arbogast *et al.*, 2002). By contrast, potential acoustical effects would be more likely to improve speech segregation by reducing *energetic* masking.

In this experiment it was noted that the benefit of spatial separation on the MVP appeared to be greater in the 'different talker' condition than in the 'same talker' condition. A contrasting result was reported recently by Noble and Perrett (2002), who found that spatial separation (in the horizontal plane) conferred a greater intelligibility advantage for same-sex talkers than different-sex talkers. This seeming discrepancy likely has an explanation in terms of the relative amounts of

informational and energetic masking release provided by separation in different dimensions. In Noble and Perrett's study, horizontal separation would have induced a large head shadow effect and allowed a strong release from *energetic* masking. This was likely the dominant effect, and hence was most apparent for similar voices (i.e. more spectral overlap). MVP separation would not confer as strong a release from energetic masking, and as such one would not expect large improvements for similar talkers. In the present study then, improvements may have been more strongly influenced by reductions in *informational* masking. In other words, the advantage of spatial separation on masking appears to be a complex function of the type(s) of masking occurring in a stimulus, and the type(s) of masking release provided by the particular change in spatial configuration.

Finally, it has been suggested that the binaural advantage becomes increasingly important with more than two talker locations for speech segregation (Yost *et al.*, 1996) and intelligibility (Hawley *et al.*, 2004). It remains to be seen whether the rather subtle 'spectral cue advantage' observed here would persist with the addition of further target locations. Presumably the extraction of spectral cues would be hindered with increasing numbers of stimuli, as specific notches and peaks became less salient. However, the percept of externalisation provided by the spectral cues (recall section 1.3.4) may still confer an advantage for selective attention in such situations (see section 8.3 for further discussion of this point).

## 7.5 Conclusions

In this experiment the effect of spatial separation of a target talker from two competing talkers was examined. Consistent with many previous studies, listeners were better able to follow the target talker with horizontal (lateral angle) separation. It was also found that vertical (polar angle) separation on the median vertical plane improved performance. This improvement was smaller than in the horizontal case, and was more pronounced when the three talkers had different voices. These findings suggest that spectral cues to sound source location can provide a modest benefit for selective listening in multiple talker situations. This benefit may arise through reductions in energetic and/or informational masking.

# Chapter 8: General discussion

## 8.1 Summary of findings

The aim of the experimental series described, as a whole, was to examine auditory spatial performance when sounds are presented from different locations concurrently. An attempt was made to examine this issue systematically by measuring spatial resolution in various spatial regions using various stimulus types presented concurrently with different relative locations.

Although different stimuli and stimulus manipulations were adopted, the general approach was to use broadband stimuli as these are localised most accurately in a single-source situation. A side-effect of this approach was that competing stimuli overlapped heavily in *both frequency and time*, providing the auditory system with a challenging task.

In general, it was found that spatial hearing was well maintained when two auditory objects were present. The auditory system appears to have some ability to extract spatial cues from mixed signals received at the ears. However, performance in spatial tasks with concurrent sources (localisation, discrimination, segregation) was not uniform across space. Significant variations in performance were observed as a function of the configuration of the competing sources with respect to each other and to the listener. The patterns observed in the results indicated that the representation of concurrent broadband sound sources is limited by the success of extraction of the relevant spatial cues. When a binaural difference was present, performance was generally good. However, for configurations in which spectral cue analysis was required, performance was far less robust, being affected by stimulus similarity, duration of the presentation, and the nature of the task.

## 8.2 Discussion of findings

### 8.2.1 Binaural cues and concurrent source perception

Binaural processing was found to be reasonably robust with two source locations. When the stimuli were separated horizontally, delivering different binaural cues, there was good evidence that two locations were perceived. For paired broadband noises, the presence of two binaural cues enabled subjects to hear two sources rather than one (Chapter 3) and to localise one source in such a pair (Chapter 4). For paired speech sources, subjects were able to localise one or both of the pair given differences in binaural position (Chapter 6). Furthermore, when the competing stimuli were ongoing sentences, binaural differences greatly improved the ability of subjects to stream the sentences and allocate simultaneous words to the appropriate sources (Chapter 7).

Two major limitations on the ability of listeners to use binaural cues in concurrent source situations were revealed. Firstly, sensitivity to binaural separation of stimulus pairs varied across space, with thresholds increasing with increasing laterality. This is in line with what is known about just-noticeable differences (JNDs) in ITD and ILD, although it seemed that separations required for concurrent spatial tasks exceeded those dictated by JNDs in the cues. Secondly, when localisation of a target sound in a concurrent pair was examined closely, a bias in perceived lateral location was observed. This bias was consistent and appeared to represent a bias away from the competing source. This effect was discussed in detail in Chapter 4, where it was proposed that it may be explained by interference in the processing of one or both of the binaural cues.

There has been little attention given to how binaural cues might be extracted from simultaneous sound sources. Patterns of ITD tuning in the auditory system are consistent with a system that is optimised to measure one ITD at a time and it has been suggested that multiple sources must be perceived by sequential estimates of each source (Fitzpatrick *et al.*, 1997). Such a mechanism is well suited to natural environments where stimuli fluctuate over time and different sources may dominate at different times. Such a mechanism, for example, can explain the ability of subjects in the experiments described here to use binaural cues effectively with competing speech sources. If such a mechanism is at work however, it must be sensitive to even very

small fluctuations in concurrent sources. For example, in Chapter 3 it was seen that ITDs could be extracted from concurrent broadband noises which differed only in their fine spectro-temporal structures. The use of short-term ITD estimates has been shown to be a robust mechanism for localising ongoing *single* sources in heavily reverberant environments (Shinn-Cunningham and Kawakyu, 2003). Thus it may be that the auditory system uses short-term ITD estimation as a primary tool for the analysis of objects in complex acoustic environments.

### 8.2.2 Spectral cues and concurrent source perception

One of the recurring questions throughout this thesis was what is the role, if any, of the spectral cues in multiple source situations? Their role in the accurate localisation of single sources is clear, and this was confirmed for speech stimuli in Chapter 5. However their role in concurrent source listening is more difficult to define. Their utility in spatial tasks was found to be strongly dependent on stimulus characteristics and the nature and complexity of the task. The clearest opportunity to examine spectral cue processing was in configurations where stimulus pairs were separated along the median vertical plane (MVP) and binaural differences were negligible. In this configuration, it was found that spectral cues were not sufficient for resolving two broadband noises with near-identical temporal envelopes (Chapter 3). However, when the stimuli differed in their temporal envelopes (Chapter 4) subjects were able to localise a target source with minor non-systematic disruptions from the distracter.

Presumably then the effective extraction of spectral cues relies on substantial amplitude fluctuations in the competing sources. It may be that spectral analysis is conducted sequentially on the different stimuli, similarly to the idea proposed above for ITD processing. If so, a relatively simple model such as that of Hofman and Van Opstal (1998) may be appropriate. In this model, very short segments of a signal are analysed. The spectrum is calculated and correlated with neurally stored HRTF representations to give an estimate of elevation. Integration then occurs over several estimates (tens of milliseconds) to give a final estimation of elevation. Hofman and Van Opstal noted that the *consistency* of the short-term estimates influences the stability of the final percept, and indeed systematic variations in the estimates may be an important factor in indicating the presence of more than one source. When pairs of

spoken words were presented on the MVP an interesting finding was made. Listeners could localise a target word with good accuracy, but were unable to localise *both* of the sources in order to discriminate their vertical positions (Chapter 6). This raised the notion that the utilisation of multiple spectral cues places high demands on processing time and/or attentional resources.

For ongoing speech stimuli, where processing time was less of an issue, it was found that separation on the MVP could aid in the segregation of concurrent talkers (Chapter 7). Although this benefit was small in comparison to the advantage of binaural separation, it showed that the different spectral cues associated with different locations can be processed and can influence performance. In addition to their role in defining different locations along the MVP (or other cones of confusion), it has been shown that the spectral cues contribute to the advantage obtained from horizontal separation. Using headphone stimulation and a competing speech paradigm, McAnally and colleagues showed that full spatialisation resulted in a significantly greater spatial improvement than stimuli separated in ITD only (McAnally *et al.*, 2003). Recent results from our laboratory have also shown that spatialisation provides a speech segregation advantage over presentations preserving ITD *and* ILD cues (Carlile *et al.*, 2004). It is likely that the percept of *externalisation* provided by the spectral cues (see section 1.3.4) is a key factor in the advantage they provide. The perception of sources in extra-personal space may increase the ability of a listener to focus attention on a target source and ignore any distracters. This was discussed in terms of reducing informational masking in Chapter 7 (section 7.4.4).

### 8.2.3 The relationship between localisation and segregation

An issue that arises when considering spatial hearing in complex environments is the relationship between localisation and segregation of objects. Some discussion was given in Chapter 4 (section 4.7.3) to this topic, where it was noted that imperfect segregation of competing sources could lead to errors in localisation. Several authors have tried to assess whether object formation must occur before localisation, or whether location may drive object formation. The present set of experiments has added mixed contributions to this issue.

In a recent imaging study, Zattore and colleagues (Zattore *et al.*, 2002) probed the so-called 'what' and 'where' pathways in human auditory cortex. Using positron emission tomography (PET), they identified a posterior region that was sensitive to the spatial distribution of sources but only when multiple complex stimuli were presented concurrently. They suggested that the 'where' pathway depends on object formation (or the 'what' pathway). This idea has a history of support in the literature, with much evidence to suggest that object formation occurs *prior to* localisation on the basis of ITD and ILD (see for example Buell and Hafter, 1991; Woods and Colburn, 1992). Psychophysical studies have also shown that spatial cues cannot be used to *drive* the segregation of concurrent vowels (Culling and Summerfield, 1995) but are useful for attending to objects that have already been grouped (Darwin and Hukin, 1999). Certainly in the present experiments there was evidence that grouping occurs before localisation. In Chapter 4, for example, it was noted that many of the errors in target localisation could be explained by incomplete segregation from maskers. This was most apparent in the 'inverse' condition of Experiment 2 (see section 4.6) where the target was highly disrupted when grouping was impaired by the presence of a dominant masker.

However, recent evidence seems to argue that spatial cues *can* drive segregation. Drennan *et al.* (2003) showed that in the free-field, realistic spatial cues *can* drive the segregation of concurrent vowels. They also went on to show that some listeners, with training, could use ITD alone for the task. In support of this, the results presented in Chapter 3 indicated that a difference in ITD could drive the resolution of two broadband sources which were otherwise undistinguishable. Thus it may be that spatial cues can drive segregation in extreme cases, where other more salient cues are perhaps unavailable. Indeed there is mounting evidence that ITD can influence segregation at some low-level before the conscious awareness of location. Patient studies have contributed intriguing pieces of evidence towards this hypothesis, including a recent investigation by Thiran and Clarke (2003). These authors examined lateralisation and segregation abilities (on the basis of ITD) in a patient with a right parieto-frontal ischaemic lesion. They reported that while lateralisation was completely abolished, segregation was still improved by separation in ITD, implicating relatively low-level binaural processing in grouping and segregation. Whether or not this role applies to the other spatial cues is not clear.

### 8.2.4 Practical relevance of results

The work described in this thesis made use of virtual auditory space (VAS) for the flexible presentation of multiple sound sources. VAS is a powerful tool, and is being employed with increasing regularity in both psychophysical and physiological research (for recent examples see Culling *et al.*, 2003; Sterbing *et al.*, 2003; Behrend *et al.*, 2004; Best *et al.*, 2004; Hawley *et al.*, 2004). Virtual acoustics also has a variety of practical applications, most notably perhaps being auditory displays for aircraft pilots (McKinley *et al.*, 1994, 1997). In this complex operating environment, the aim is to convey information via sound in order to reduce overcrowding of the important visual channel. Another important application, where the auditory system is relied on for the delivery of accurate spatial information, is in navigational aids for the blind or those with obstructed vision (such as fire-fighters in heavy smoke). In all of these situations, spatial audio is a meaningful way to indicate the position of objects in the environment, and it is also a convenient communication channel for the delivery of warnings and dialogue.

An interesting interaction has arisen between psychophysical research and virtual space technology. While VAS has had a powerful impact on psychophysics, psychophysics is also now having an impact on the development and refinement of VAS. For instance, it may be a requirement that a 3D audio display be able to deliver multiple auditory objects whilst maintaining a clear percept of the individual signals and their locations. For this implementation to be a success, it is important that basic capabilities of the human auditory system for localising and discriminating concurrent sources be kept in mind.

A simple implementation might involve the delivery of auditory tokens to a listener to convey spatially referenced information or alerts. If there is a chance that these tokens will occur simultaneously, it is clear from the experiments described in Chapters 3-6 that spatial layout will impact on the fidelity of the spatial percept. Requirements for spatial layout will depend on the spectral and temporal properties of the competing signals, and the specific tasks required (e.g. detection, localisation, or discrimination). Spatial layout is also an important consideration in more complex auditory displays with information-carrying streams of speech arising from multiple talkers. In a recent study concerned with presenting up to seven competing talkers in a

spatial auditory display, Brungart and colleagues noted that the azimuthal variation in spatial sensitivity to speech must be kept in mind when optimally spacing talkers (Brungart and Simpson, 2003). The experiment described in Chapter 7 added to this by giving some insight into the utility of different kinds of spatial cues for speech segregation. As a whole, it is clear that an ongoing deepening of our understanding and characterisation of spatial hearing will be crucial for the development of increasingly sophisticated virtual environments.

## 8.3 Key areas for further research

As has been emphasised throughout this thesis, complex environments containing concurrent sound sources present interesting challenges to the auditory system, and particularly to the processes underlying spatial hearing. The psychophysical experiments described in this thesis and those of other authors provide us with many clues as to how the nervous system might represent multiple sources. However, complementary physiological investigations are vital if we are to test the validity of these speculations. The appearance of several recent publications provides encouragement that advances in neurophysiology are underway in this area (see next section). In terms of behavioural research, there are many directions that await exploration, and just some of these are touched on below.

### 8.3.1 Neurophysiology

Auditory neurophysiologists have contributed an enormous amount to our current understanding of the mechanisms for sound localisation, and just a fraction of this work was discussed in Chapter 1. However, a relatively small number of studies have investigated situations involving multiple concurrent sound sources. Those studies that have investigated concurrent source processing in the animal brain have provided a good foundation for further work in this area. Takahashi and Keller, working with a well described neural spatial map in the barn owl inferior colliculus, presented broadband uncorrelated stimulus pairs very similar to those described in Chapter 3 of this thesis (Takahashi and Keller, 1994). They observed that single neurons were able to respond to the two sources separately (when they were separated by 45° in

azimuth), indicating that they were resolved in the spatial map. They also observed that activity was reduced compared to single source responses, and questioned whether this may cause behavioural localisation deficits when multiple sources are present. More recently, in the cat inferior colliculus, it was reported that changes in activity of low-frequency neurons occur systematically with separation of competing stimuli, and this was identified as a correlate of the well known spatial release from masking phenomenon (Lane *et al.*, 2003).

Some work has also been done in cat auditory cortex recently, although primarily focussing on the neural responses to pairs of stimuli with non-simultaneous onsets (such as a source and its delayed echo, see Reale and Brugge, 2000; Mickey and Middlebrooks, 2001). One study did investigate the impact of a masking noise source on responses of neuron clusters in area A2 to click-train target stimuli (Furukawa and Middlebrooks, 2001). They found that response rates and latencies were altered by the masker and these alterations were greatest for coincident locations. Although many alterations were equivalent to those induced by a simple reduction in target level, many others did not fit this description. The authors concluded that some features of cortical spike patterns are disrupted by the presence of a masker, whereas others are not. This is an interesting finding in light of the results of this thesis (Chapter 4; Chapter 6 Experiment 2), where gross localisation was unaffected by a masker, but fine disturbances were observed. What remains to be elucidated is what the coding scheme is, and just how multiple sources are represented. This question is particularly complex because a simple spatiotopic map is not present in auditory cortex (recall section 1.3.5) and large populations of neurons appear to be involved in the representation of a single location. Further studies are needed to shed light on the issue of multiple-source coding schemes, and these might well attempt to extend existing models of single-source cortical coding (Brugge *et al.*, 1996; Middlebrooks *et al.*, 1994; Middlebrooks *et al.*, 1998).

Whether future physiological investigations concentrate on the inferior colliculus, auditory cortex or other areas of the brain, there are several interesting questions to pursue. It would be interesting, for example, to examine the time course for the processing of single and multiple auditory objects. Certainly some of the results of this thesis indicated that concurrent source analysis takes longer than would be predicted by single source integration times (see section 6.4). Another intriguing question concerns the effect of attention on spatial representations in the auditory

system. Many studies have showed that activity in different auditory centres is altered by attention (just a few examples include monkey cortex: Benson and Heinz, 1978; cat superior colliculus: Meredith and Clemo, 1989; and human cortex: Woldorff and Hillyard, 1991; Grady *et al.*, 1997). Furthermore, many descending pathways have been identified that may underlie attentional effects (Huffman and Henson, 1990; Spangler and Warr, 1991). As our understanding of spatial representations in the auditory system increases, an opportunity arises to investigate how descending inputs might alter these representations when attention is directed spatially (see next section).

### 8.3.2 Behavioural research

In addition to its fundamental role in communication, probably one of the most important roles of the auditory system is to monitor the environment and specifically to detect changes in the environment. Sensory change is an important concept as it generally corresponds to new objects or new situations that may require a new behaviour, and acoustic changes are particularly useful because they are detected extremely quickly (Woodworth and Schlosberg, 1954; Sanders, 1998). Auditory information can be obtained from any direction at any time (i.e. beyond the visual field) and of course is available in darkness. Moreover, the work in this thesis has confirmed that it can be obtained simultaneously from more than one direction and generally with good accuracy.

Thus in terms of perceiving the spatial layout of the environment, auditory information is an extremely important complement to visual information. No doubt the most important use of a spatial representation is the appropriate distribution of attention. The study of auditory attention has a relatively long history (e.g. see Broadbent, 1954), although it has never achieved a presence in the literature as dominant as that of visual attention. It is clear that attention can be directed spatially in audition. Attention can be directed to a particular hemifield (Spence and Driver, 1994), ITD (Sach *et al.*, 2000) or azimuthal location (Mondor and Zattore, 1995). It is also generally agreed that attention is distributed as a spatial gradient, where resources decline with distance from the focal point (Mondor and Zattore, 1995; Teder and Naatanen, 1994; Teder-Sälejärvi *et al.*, 1999). In terms of the dynamics of auditory attention, it was suggested by one author that shifts in spatial attention require an

amount of time that is proportional to the size of the spatial shift (Rhodes, 1987). However, this view has been challenged and more rigorous experiments have indicated that the time required to shift attention is *not* distance-dependent (Mondor and Zattore, 1995).

These studies have begun to build a picture of the relationship between attention and auditory spatial layout, but the picture is far from complete. The experiments described in this thesis raised several issues related to attention that have been overlooked so far. For instance, previous studies have generally examined only lateral (binaural) effects and have not considered auditory space as a 3-dimensional continuum. Chapter 7 of this thesis gave some preliminary indications that attention can also be directed differentially to locations within a binaural interval (i.e. along a cone of confusion) but this notion has yet to be extensively tested. Another point that is not yet clear is whether auditory attention can be *divided* between several locations at once (see also section 6.4), a phenomenon that there is some evidence for in the visual system (Müller *et al.*, 2003). One recent study approached this issue using multiple speech streams (Shinn-Cunningham *et al.*, 2004) and came to an interesting conclusion. The results suggested that when listeners are required to divide their attention between two locations, they employ a *selective* listening strategy to enhance the perception of the less salient (e.g. quieter) sound source.

It is hoped that further efforts to characterise auditory spatial attention will clarify some of these issues and ultimately increase our understanding of the role of auditory spatial representations in complex natural environments.

# Bibliography


Abouchacra, K. S., Emanuel, D. C., Blood, I. M., and Letowski, T. R. (1998). "Spatial perception of speech in various signal to noise ratios," *Ear and Hearing,* 19:298-309.

Algazi, V. R., Avendano, C., and Duda, R. O. (2001). "Elevation localization and head-related transfer function analysis at low frequencies," *J. Acoust. Soc. Am.,* 109:1110-1122.

Angell, J. R. and Fite, W. (1901). "The monaural localization of sound," *Psychol. Rev.,* 8:225-243.

Arbogast, T. L., Mason, C. R., and Kidd, G. (2002). "The effect of spatial separation on informational and energetic masking of speech," *J. Acoust. Soc. Am.,* 112:2086-2098.

Aungst, J. L., Heyward, P. M., Puche, A. C., Karnup, S. V., Hayar, A., Szabo, G., and Shipley, M. T. (2003). "Centre surround inhibition among olfactory bulb glomeruli," *Nature,* 426:623-629.

Begault, D. R. and Wenzel, E. M. (1993). "Headphone localization of speech," *Human Factors,* 35:361-376.

Begault, D. R., Wenzel, E. M., and Anderson, M. R. (2001). "Direct comparison of the impact of head-tracking, reverberation, and individualized head-related transfer functions on the spatial perception of a virtual speech source," *J. Audio Eng. Soc.,* 49:904-916.

Behrend, O., Clarke, E., Dickson, B., and Carlile, S. (2004). "Location-dependent 1st spike latency in the guinea pig midbrain in virtual acoustic space and free field," in *Proc. Australian Neuroscience Soc.*, 15, Melbourne, Australia, 93.

Benson, D. A. and Heinz, R. D. (1978). "Single-unit activity in the auditory cortex of monkeys selectively attending left vs. right ear stimuli," *Brain Research,* 159:307-320.

Bernstein, L. R. and Trahiotis, C. (1982). "Detection of interaural delay in high-frequency noise," *J. Acoust. Soc. Am.,* 71:147-152.

Best, V., van Schaik, A., and Carlile, S. (2002). "The perception of multiple broadband noise sources presented concurrently in virtual auditory space," in *Proc. of the Audio Eng. Soc. 112th Convention*, Munich, Germany, paper 5549.

Best, V., van Schaik, A., and Carlile, S. (2004). "Separation of concurrent broadband sound sources by human listeners," *J. Acoust. Soc. Am.,* 115:324-336.

Binns, K. E., Grant, S., Withington, D. J., and Keating, M. J. (1992). "A topographic representation of auditory space in the external nucleus of the inferior colliculus of the guinea-pig," *Brain Res.,* 589:231-242.

Blauert, J. (1983). *Spatial Hearing: The Psychophysics of Human Sound Localization*, MIT Press, Cambridge.

Blauert, J. and Lindemann, W. (1986). "Spatial mapping of intracranial auditory events for various degrees of interaural coherence," *J. Acoust. Soc. Am.,* 79:806-813.

Bolia, R. S., Nelson, W. T., Ericson, M. A., and Simpson, B. D. (2000). "A speech corpus for multitalker communications research," *J. Acoust. Soc. Am.,* 107:1065-1066.

Botte, M. C., Drake, C., Brochard, R., and McAdams, S. (1997). "Perceptual attenuation of nonfocused auditory streams," *Percept. Psychophys.,* 59:419-425.

Braasch, J. (2002). "Localization in the presence of a distracter and reverberation in the frontal horizontal plane: II. Model algorithms," *Acta acustica/Acustica,* 88:956-969.

Braasch, J. and Hartung, K. (2002). "Localization in the presence of a distracter and reverberation in the frontal horizontal plane: I. Psychoacoustical data," *Acta acustica/Acustica,* 88:942-955.

Brand, A., Behrend, O., Marquardt, T., McAlpine, D., and Groethe, B. (2002). "Precise inhibition is essential for microsecond interaural time difference coding," *Nature,* 417:543-547.

Bregman, A. S. (1990). *Auditory Scene Analysis: The Perceptual Organization of Sound*, MIT Press, Cambridge.

Bridgemann, B., AIken, W., Allen, J., and Maresh, T. C. (1997). "Influence of acoustic context on sound localization: An auditory Roelofs effect," *Psychol. Res.,* 60:238-243.

Broadbent, D. E. (1954). "The role of auditory localization in attention and memory span," *J. Exp. Psychol.,* 47:191-196.

Brokx, J. P. L. and Nooteboom, S. G. (1982). "Intonation and the perceptual separation of simultaneous voices," *J. Phonetics,* 10:23-36.

Bronkhorst, A. W. (1995). "Localization of real and virtual sound sources," *J. Acoust. Soc. Am.,* 98:2542-2553.

Bronkhorst, A. W. and Plomp, R. (1988). "The effect of head-induced interaural time and level differences on speech intelligibility in noise," *J. Acoust. Soc. Am.,* 83:1508-1516.

Brown, C. H., Schessler, T., Moody, D., and Stebbins, W. (1982). "Vertical and horizontal sound localization in primates," *J. Acoust. Soc. Am.,* 72:1804-1811.

Brugge, J. F., Reale, R. A., and Hind, J. E. (1996). "The structure of spatial receptive fields of neurons in primary auditory cortex of the cat," *J. Neurosci.,* 16:4420-4437.

Brungart, D. S. and Rabinowitz, W. R. (1999). "Auditory localization of nearby sources. Head-related transfer functions," *J. Acoust. Soc. Am.,* 106:1465-1479.

Brungart, D. S. and Simpson, B. D. (2002). "The effects of spatial separation in distance on the informational and energetic masking of a nearby speech signal," *J. Acoust. Soc. Am.,* 112:664-676.

Brungart, D. S. and Simpson, B. D. (2003). "Optimizing the spatial configuration of a seven-talker speech display," in *Proc. Int. Conf. Auditory Display*, Boston, USA, 188-191.

Brungart, D. S., Simpson, B. D., Ericson, M. A., and Scott, K. R. (2001). "Informational and energetic masking effects on the perception of multiple simultaneous talkers," *J. Acoust. Soc. Am.,* 110:2527-2538.

Buell, T. N. and Hafter, E. R. (1991). "Combination of binaural information across frequency bands," *J. Acoust. Soc. Am.,* 90:1894-1900.

Buell, T. N., Trahiotis, C., and Bernstein, L. R. (1991). "Lateralization of low-frequency tones: Relative potency of gating and ongoing interaural delays," *J. Acoust. Soc. Am.,* 90:3077-3085.

Butler, R. A. (1986). "The bandwidth effect on monaural and binaural localization," *Hearing Res.,* 21:67-73.

Butler, R. A. and Belendiuk, K. (1977). "Spectral cues utilized in the localization of sound in the median sagittal plane," *J. Acoust. Soc. Am.,* 61:1264-1269.

Butler, R. A., Humanski, R. A., and Musicant, A. D. (1990). "Binaural and monaural localization of sound in two-dimensional space," *Perception,* 19:241-256.

Butler, R. A. and Naunton, R. F. (1964). "Role of stimulus frequency and duration in the phenomenon of localization shifts," *J. Acoust. Soc. Am.,* 36:917-922.

Butler, R. L. and Humanski, R. A. (1992). "Localization of sound in the vertical plane with and without high-frequency spectral cues," *Percept. Psychophys.,* 51:182-186.

Caird, D. and Klinke, R. (1987). "Processing of interaural time and intensity differences in the cat inferior colliculus," *Exp. Brain Res.,* 68:379-392.

Canévet, G. and Meunier, S. (1996). "Effect of adaptation on auditory localization and lateralization," *Acta acustica/Acustica,* 82:149-157.

Carlile, S. (1996a). "The physical and psychophysical basis of sound localization," in *Virtual Auditory Space: Generation and Applications,* ed. S. Carlile, Landes, Austin, Chapter 2.

Carlile, S. (1996b). *Virtual Auditory Space: Generation and Applications*, Landes, Austin.

Carlile, S., Al-Mahaidi, S., and Leung, J. (2004). "Spatialisation of talkers and the segregation of concurrent speech," in *Proc. Assoc. Res. Otolaryng. 27th Mid-winter Meeting*, Daytona, USA, 209-210.

Carlile, S. and Delaney, S. (1999). "The localization of spectrally restricted sounds by human listeners," *Hearing Res.,* 128:175-189.

Carlile, S., Hyams, S., and Delaney, S. (2001). "Systematic distortions of auditory space perception following prolonged exposure to broadband noise," *J. Acoust. Soc. Am.,* 110:416-424.

Carlile, S., Jin, C., and van Raad, V. (2000). "Continuous virtual auditory space using HRTF interpolation: Acoustic and psychophysical errors," in *Proc. First IEEE Pacific-Rim Conf. on Multimedia*, Sydney, Australia, 220-223.

Carlile, S. and King, A. J. (1994). "Monaural and binaural spectrum level cues in the ferret: Acoustics and the neural representation of auditory space," *J. Neurophysiol.,* 71:785-801.

Carlile, S., Leong, P., and Hyams, S. (1997). "The nature and distribution of errors in sound localization by human listeners," *Hearing Res.,* 114:179-196.

Carlile, S. and Leung, J. (2001). "Rendering sound sources in high fidelity virtual auditory space: Some spatial sampling and psychophysical factors," in *Usability Evaluation and Interface Design: Cognitive Engineering, Intelligent Agents and Virtual Reality,* ed. M. Smith, G. Salvendy, D. Harris and R. Koubek, Lawrence Erlbaum, New Jersey, 599-603.

Carlile, S. and Pettigrew, A. G. (1987). "Distribution of frequency sensitivity in the superior colliculus of the guinea pig," *Hearing Res.,* 31:123-136.

Carlile, S. and Pralong, D. (1994). "The location-dependent nature of perceptually salient features of the human head-related transfer function," *J. Acoust. Soc. Am.,* 95:3445-3459.

Carr, C. E. and Konishi, M. (1990). "A circuit for detection of interaural time differences in the brain stem of the barn owl," *J. Neurosci.,* 10:3227-3246.

Casseday, J. H. and Neff, W. D. (1975). "Auditory localization: Role of auditory pathways in brain stem of the cat," *J. Neurophysiol.,* 38:842-858.

Colburn, H. S. (1996). "Computational models of binaural processing," in *Auditory Computation,* ed. H. L. Hawkins, T. A. McMullen, A. N. Popper and R. R. Fay, Springer-Verlag, New York,

Culling, J. F., Hodder, K. I., and Toh, C. Y. (2003). "Effects of reverberation on perceptual segregation of competing voices," *J. Acoust. Soc. Am.,* 114:2871-2876.

Culling, J. F. and Summerfield, Q. (1995). "Perceptual separation of concurrent speech sounds: Absence of across-frequency grouping by common interaural delay," *J. Acoust. Soc. Am.,* 98:785-797.

Damaske, P. (1967/68). "Subjektive Untersuchungen von Schallfeldern [Subjective investigations of sound fields]," *Acustica,* 19:198-213.

Darwin, C. J. and Carlyon, R. P. (1995). "Auditory grouping," in *The Handbook of Perception and Cognition, Volume 6, Hearing,* ed. B. C. J. Moore, Academic Press, London, 387-424.

Darwin, C. J. and Hukin, R. W. (1999). "Auditory objects of attention: The role of interaural time differences," *J. Exp. Psychol.,* 25:617-629.

Darwin, C. J. and Hukin, R. W. (2000). "Effectiveness of spatial cues, prosody, and talker characteristics in selective attention," *J. Acoust. Soc. Am.,* 107:970-977.

Davis, K. A., Ramachandran, R., and May, B. J. (2003). "Auditory processing of spectral cues for sound localization in the inferior colliculus," *J. Assoc. Res. Otolaryngology,* 4:148-163.

Delgutte, B., Joris, P. X., Litovsky, R. Y., and Yin, T. C. T. (1995). "Relative importance of different acoustic cues to the directional sensitivity of inferior-colliculus neurons," in *Advances in Hearing Research - Proceedings of the*

*10th International Symposium on Hearing,* ed. G. A. Manley, G. M. Klump, C. Köppl, H. Fastl and H. Oeckinghaus, World Scientific, London, 288-297.

Delgutte, B., Joris, P. X., Litovsky, R. Y., and Yin, T. C. T. (1999). "Receptive fields and binaural interactions for virtual-space stimuli in the cat inferior colliculus," *J. Neurophysiol.,* 81:2833-2851.

Dirks, D. D. and Wilson, R. H. (1969). "The effect of spatially separated sound sources on speech intelligibility," *J. Speech Hear. Res.,* 12:5-38.

Divenyi, P. L. and Oliver, S. K. (1989). "Resolution of steady-state sounds in simulated auditory space," *J. Acoust. Soc. Am.,* 85:2042-2052.

Domnitz, R. H. and Colburn, H. S. (1977). "Lateral position and interaural discrimination," *J. Acoust. Soc. Am.,* 61:1586-1598.

Drennan, W. R., Gatehouse, S. G., and Lever, C. (2003). "Perceptual segregation of competing speech sounds: The role of spatial location," *J. Acoust. Soc. Am.,* 114:2178-2189.

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

Drullman, R. and Bronkhorst, A. W. (2000). "Multichannel speech intelligibility and talker recognition using monaural, binaural, and three-dimensional auditory presentation," *J. Acoust. Soc. Am.,* 107:2224-2235.

Durlach, N. I., Mason, C. R., Kidd, G., Arbogast, T. L., Colburn, H. S., and Shinn-Cunningham, B. G. (2003). "Note on informational masking," *J. Acoust. Soc. Am.,* 113:2984-2987.

Dye, R. H. (1990). "The combination of interaural information across frequencies: Lateralization on the basis of interaural delay," *J. Acoust. Soc. Am.,* 88:2159-2170.

Dye, R. H., Yost, W. A., Stellmack, M. A., and Sheft, S. (1994). "Stimulus classification procedure for assessing the extent to which binaural processing is spectrally analytic or synthetic," *J. Acoust. Soc. Am.,* 96:2720-2730.

Egan, J. P. (1948). "Articulation testing methods," *Laryngoscope,* 58:955-991.

Finney, D. J. (1971). *Probit Analysis,* Cambridge University Press, Cambridge.

Fisher, N. I., Lewis, T., and Embleton, B. J. J. (1987). *Statistical Analysis of Spherical Data*, Cambridge University Press, Cambridge.

Fitzpatrick, D. C., Batra, R., Stanford, T. R., and Kuwada, S. (1997). "A neuronal population code for sound localization," *Nature,* 388:871-874.

Fitzpatrick, D. C., Kuwada, S., and Batra, R. (2000). "Neural sensitivity to interaural time differences: Beyond the Jeffress model," *J. Neurosci.,* 20:1605-1615.

Freyman, R. L., Balakrishnan, U., and Helfer, K. S. (2001). "Spatial release from informational masking in speech recognition," *J. Acoust. Soc. Am.,* 109:2112-2122.

Freyman, R. L., Helfer, K. S., McCall, D. D., and Clifton, R. K. (1999). "The role of perceived spatial separation in the unmasking of speech," *J. Acoust. Soc. Am.,* 106:3578-3588.

Fujiki, N., Riederer, K. A. J., Jousmäki, V., Mäkelä, J. P., and Hari, R. (2002). "Human cortical representation of virtual auditory space: Differences between sound azimuth and elevation," *European J. Neurosci.,* 16:2207-2213.

Furukawa, S. and Middlebrooks, J. C. (2001). "Sensitivity of auditory cortical neurons to locations of signals and competing noise sources," *J. Neurophysiol.,* 86:226-240.

Gabriel, K. J. and Colburn, H. S. (1981). "Interaural correlation discrimination: I. Bandwidth and level dependence," *J. Acoust. Soc. Am.,* 69:1394-1401.

Gaik, W. (1993). "Combined evaluation of interaural time and intensity differences: Psychoacoustic results and computer modeling," *J. Acoust. Soc. Am.,* 94:98-110.

Gardner, M. B. and Gardner, R. S. (1973). "Problem of localization in the median plane: Effect of pinnae cavity occlusion," *J. Acoust. Soc. Am.,* 53:400-408.

Getzmann, S. (2002). "The effect of eye position and background noise on vertical sound localization," *Hearing Res.,* 169:130-139.

Getzmann, S. (2003a). "A comparison of the contrast effects in sound localization in the horizontal and vertical planes," *Exp. Psychol.,* 50:131-141.

Getzmann, S. (2003b). "The influence of acoustic context on vertical sound localisation in the median plane," *Percept. Psychophys.,* 65:1045-1057.

Gilkey, R. H. and Anderson, T. R. (1995). "The accuracy of absolute localization judgments for speech stimuli," *J. Vestib. Res.,* 5:487-497.

Gilkey, R. H. and Good, M. D. (1995). "Effects of frequency on free-field masking," *Human Factors,* 37:835-843.

Golay, M. J. E. (1961). "Complementary series," *IRE Trans. Inf. Theory,* 7:82-87.

Goldberg, J. M. and Brown, P. B. (1969). "Response of binaural neurons of dog superior olivary complex to dichotic tonal stimuli: Some physiological mechanisms of sound localization," *J. Neurophysiol.,* 32:613-636.

Good, M. D. and Gilkey, R. H. (1996). "Sound localization in noise: The effect of signal-to-noise ratio," *J. Acoust. Soc. Am.,* 99:1108-1117.

Good, M. D., Gilkey, R. H., and Ball, J. M. (1997). "The relation between detection in noise and localization in noise in the free field," in *Binaural and Spatial Hearing in Real and Virtual Environments,* ed. R. H. Gilkey and T. R. Anderson, Erlbaum, Hillsdale, NJ, 349-376.

Grady, C. L., Meter, J. W. V., Maisog, J. M., Pietrini, P., Krasuski, J., and Rauschecker, J. P. (1997). "Attention-related modulation of activity in primary and secondary auditory cortex," *NeuroReport,* 8:2511-2516.

Hafter, E. R. and Maio, J. D. (1975). "Difference thresholds for interaural delay," *J. Acoust. Soc. Am.,* 57:181-187.

Harris, J. D. (1972). "A florilegium of experiments on directional hearing," *Acta Oto-laryng. Supp.,* 298:5-26.

Harris, L. R., Blakemore, C., and Donaghy, M. (1980). "Integration of visual and auditory space in the mammalian superior colliculus," *Nature,* 288:56-59.

Harris, S. (1998). "The effect of sound level on sound localisation accuracy," Honours thesis, Department of Physiology, University of Sydney.

Hartmann, W. M. (1983). "Localization of sound in rooms," *J. Acoust. Soc. Am.,* 74:1380-1391.

Hartmann, W. M. and Wittenberg, A. (1996). "On the externalization of sound images," *J. Acoust. Soc. Am.,* 99:3678-3688.

Hartung, K., Braasch, J., and Sterbing, S. J. (1999). "Comparison of different interpolation methods for the interpolation of head-related transfer functions," in *Proc. AES 16th Int. Conf. on Spatial Sound Reproduction*, Rovaniemi, Finland, 319-329.

Hawley, M. L., Litovsky, R. Y., and Colburn, H. S. (1999). "Speech intelligibility and localization in a multi-source environment," *J. Acoust. Soc. Am.,* 105:3436-3448.

Hawley, M. L., Litovsky, R. Y., and Culling, J. F. (2004). "The benefit of binaural hearing in a cocktail party: Effect of location and type of interferer," *J. Acoust. Soc. Am.,* 115:833-843.

Hebrank, J. and Wright, D. (1974). "Spectral cues used in the localization of sound sources on the median plane," *J. Acoust. Soc. Am.,* 56:1829-1834.

Heller, L. M. and Trahiotis, C. (1996). "Extents of laterality and binaural interference effects," *J. Acoust. Soc. Am.,* 99:3632-3637.

Henning, G. B. (1974). "Detectability of interaural delay in high-frequency complex waveforms," *J. Acoust. Soc. Am.,* 55:84-90.

Henning, G. B. (1980). "Some observations on the lateralization of complex waveforms," *J. Acoust. Soc. Am.,* 68:446-454.

Hershkowitz, R. M. and Durlach, N. I. (1969). "Interaural time and amplitude jnds for a 500-Hz tone," *J. Acoust. Soc. Am.,* 46:1464-1467.

Hofman, P. M. and Van Opstal, A. J. (1998). "Spectro-temporal factors in two-dimensional human sound localization," *J. Acoust. Soc. Am.,* 103:2634-2648.

Hofman, P. M. and Van Opstal, A. J. (2003). "Binaural weighting of pinna cues in human sound localization," *Exp. Brain Res.,* 148:458-470.

Hofman, P. M., Van Riswick, J. G. A., and Van Opstal, A. J. (1998). "Relearning sound localization with new ears," *Nature Neurosci.,* 1:417-421.

Huffman, R. F. and Henson, O. W. (1990). "The descending auditory pathway and acousticomotor systems: Connections with the inferior colliculus," *Brain Res. Rev.,* 15:295-323.

Hukin, R. W. and Darwin, C. J. (1995). "Effects of contralateral presentation and of interaural time differences in segregating a harmonic from a vowel," *J. Acoust. Soc. Am.,* 98:785-797.

Humanski, R. A. and Butler, R. A. (1988). "The contribution of the near and far ear toward localization of sound in the sagittal plane," *J. Acoust. Soc. Am.,* 83:2300-2310.

Imig, T. J., Poirier, P., Irons, W. A., and Samson, F. K. (1997). "Monaural contrast mechanism for neural sensitivity to sound direction in the medial geniculate body of the cat," *J. Neurophysiol.,* 78:2754-2771.

Irvine, D. R. F. (1992). "Physiology of the auditory brainstem," in *The Mammalian Auditory Pathway: Neurophysiology,* ed. A. N. Popper and R. R. Fay, Springer-Verlag, New York, 153-231.

Jeffress, L. A. (1948). "A place theory of sound localization," *J. Comp. Physiol. Psychol.,* 41:35-30.

Jin, C. (2001). "Spectral Analysis and Resolving Spectral Ambiguities in Human Sound Localization," PhD thesis, Electrical and Information Engineering, University of Sydney.

Joris, P. X., Carney, L. H., Smmith, P. H., and Yin, T. C. (1994). "Enhancement of neural synchronization in the anteroventral cochlear nucleus. I. Responses to tones at the characteristic frequency," *J. Neurophysiol.,* 71:1022-1036.

Karlsen, B. L. (1999). "Spatial Localization of Speech Segments," PhD thesis, Aalborg University.

Kashino, M. and Nishida, S. (1998). "Adaptation in the processing of interaural time differences revealed by the auditory localization aftereffect," *J. Acoust. Soc. Am.,* 103:3597-3604.

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

Kidd, G., Mason, C. R., Rohtla, T. L., and Deliwala, P. S. (1998). "Release from masking due to spatial separation of sources in the identification of nonspeech auditory patterns," *J. Acoust. Soc. Am.,* 104:422-431.

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

King, A. J., Jiang, Z. D., and Moore, D. R. (1998). "Auditory brainstem projections to the ferret superior colliculus: Anatomical contribution to the neural coding of sound azimuth," *J. Comp. Neurol.,* 390:342-365.

King, R. B. and Oldfield, S. R. (1997). "The impact of signal bandwidth on auditory localization: implications for the design of three-dimensional auditory displays," *Human Factors,* 39:287-295.

Kistler, D. J. and Wightman, F. L. (1992). "A model of head-related transfer functions based on principal components analysis and minimum-phase reconstruction," *J. Acoust. Soc. Am.,* 91:1637-1647.

Klumpp, R. G. and Eady, H. R. (1956). "Some measurements of interaural time difference thresholds," *J. Acoust. Soc. Am.,* 28:859-860.

Kuhn, G. F. (1987). "Physical acoustics and measurements pertaining to directional hearing," in *Directional Hearing,* ed. W. A. Yost and G. Gourevitch, Springer-Verlag, New York, 3-25.

Kulkarni, A. and Colburn, H. (1998). "Role of spectral detail in sound-source localization," *Nature,* 396:747-749.

Kunov, H. and Abel, S. M. (1981). "Effects of rise/decay time on the lateralization of interaurally delayed 1-kHz tones," *J. Acoust. Soc. Am.,* 69:769-773.

Lane, C. C., Delgutte, B., and Colburn, H. S. (2003). "A population of ITD-sensitive units in the cat inferior colliculus shows correlates of spatial release from masking," in *Assoc. Res. Otolaryngology Mid-winter Meeting*, 26, Daytona Beach, Florida, 178-179.

Langendijk, E. H., Kistler, D. J., and Wightman, F. L. (2001). "Sound localization in the presence of one or two distracters," *J. Acoust. Soc. Am.,* 109:2123-2134.

Langendijk, E. H. A. and Bronkhorst, A. W. (2002). "Contribution of spectral cues to human sound localization," *J. Acoust. Soc. Am.,* 112:1583-1596.

Leong, P. and Carlile, S. (1998). "Methods for spherical data analysis and visualization," *J. Neurosci. Methods,* 80:191-200.

Lewald, J. (1998). "The effect of gaze eccentricity on perceived sound direction and its relation to visual localization," *Hearing Res.,* 115:206-216.

Lewald, J. (2002). "Vertical sound localization in blind humans," *Neuropsychologia,* 40:1868-1872.

Litovsky, R. Y., Colburn, H. S., Yost, W. A., and Guzman, S. J. (1999). "The precedence effect," *J. Acoust. Soc. Am.,* 106:1633-1654.

Lorenzi, C., Gatehouse, S., and Lever, C. (1999). "Sound localization in noise in normal-hearing listeners," *J. Acoust. Soc. Am.,* 105:1810-1820.

Macpherson, E. A. and Middlebrooks, J. C. (2002). "Listener weighting of cues for lateral angle: The duplex theory of sound localization revisited," *J. Acoust. Soc. Am.,* 111:2219-2236.

Macpherson, E. A. and Middlebrooks, J. C. (2003). "Vertical-plane sound localization probed with ripple-spectrum noise," *J. Acoust. Soc. Am.,* 114:430-445.

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

May, B. J. (2000). "Role of the dorsal cochlear nucleus in the sound localization behavior of cats," *Hearing Res.,* 148:74-87.

McAlpine, D., Jiang, D., and Palmer, A. R. (2001). "A neural code for low-frequency sound localisation in mammals," *Nature Neurosci.,* 4:396-401.

McAnally, K., Martin, R., Bolia, R., and Brungart, D. (2003). "The use of virtual audio for the spatial segregation of competing speech," in *Proc. 8th Western Pacific Acoust. Conf.*, Melbourne, Australia, TE32.

McFadden, D. and Pasanen, E. G. (1976). "Lateralization at high frequencies based on interaural time differences," *J. Acoust. Soc. Am.,* 59:634-639.

McKinley, R. L., Erickson, M. A., and D'Angelo, W. R. (1994). "3-dimensional auditory displays: Development, applications, and performance," *Aviat. Space Environ. Med.,* 65:A31-8.

McKinley, R. L., Erickson, M. A., and D'Angelo, W. R. (1997). "Flight demonstration of a 3-D auditory display," in *Binaural and Spatial Hearing in Real and Virtual Environments,* ed. R. H. Gilkey and T. R. Anderson, Lawrence Erlbaum, New Jersey, 683–699.

Meredith, M. A. and Clemo, H. R. (1989). "Auditory cortical projection from the anterior ectosylvian sulcus (field AES) to the superioir colliculus in the cat: An anatomical and electrophysiological study," *J. Comp. Neurol.,* 289:687-707.

Mershon, D. H. and Bowers, J. N. (1979). "Absolute and relative cues for the auditory perception of egocentric distance," *Perception,* 8:311-322.

Mershon, D. H. and King, L. E. (1975). "Intensity and reverberation as factors in the auditory perception of egocentric distance," *Percept. Psychophys.,* 18:409-415.

Mickey, B. J. and Middlebrooks, J. C. (2001). "Responses of auditory cortical neurons to pairs of sounds: Correlates of fusion and localization," *J. Neurophysiol.,* 86:1333-1350.

Middlebrooks, J. C. (1992). "Narrow-band sound localization related to external ear acoustics," *J. Acoust. Soc. Am.,* 92:2607-2624.

Middlebrooks, J. C. (1999a). "Individual differences in external-ear transfer functions reduced by scaling in frequency," *J. Acoust. Soc. Am.,* 106:1480-1492.

Middlebrooks, J. C. (1999b). "Virtual localization improved by scaling non-individualized external-ear transfer functions in frequency," *J. Acoust. Soc. Am.,* 106:1493-1509.

Middlebrooks, J. C., Clock, A. E., Xu, L., and Green, D. M. (1994). "A panoramic code for sound location by cortical neurons," *Science,* 264:842-844.

Middlebrooks, J. C. and Green, D. M. (1990). "Directional dependence of interaural envelope delays," *J. Acoust. Soc. Am.,* 87:2149-2162.

Middlebrooks, J. C. and Green, D. M. (1991). "Sound localization by human listeners," *Annu. Rev. Psychol.,* 42:135-159.

Middlebrooks, J. C. and Knudsen, E. I. (1984). "A neural code for auditory space in the cat's superior colliculus," *J. Neurosci.,* 4:2621-2634.

Middlebrooks, J. C., Makous, J. C., and Green, D. M. (1989). "Directional sensitivity of sound pressure levels in the human ear canal," *J. Acoust. Soc. Am.,* 86:89-108.

Middlebrooks, J. C., Xu, L., Eddins, A. C., and Green, D. M. (1998). "Codes for sound-source location in nontonotopic auditory cortex," *J. Neurophysiol.,* 80:863-881.

Mills, A. W. (1958). "On the minimum audible angle," *J. Acoust. Soc. Am.,* 30:237-246.

Mills, A. W. (1972). "Auditory localization," in *Foundations of Modern Auditory Theory,* ed. J. V. Tobias, Academic Press, New York, 303-348.

Møller, H., Sørensen, M. F., Hammershøi, D., and Jensen, C. B. (1995). "Head-related transfer functions of human subjects," *J. Audio Eng. Soc.,* 43:300-321.

Mondor, T. A. and Zattore, R. J. (1995). "Shifting and focusing auditory spatial attention," *J. Exp. Psych. Human Percept. Perform.,* 21:387-409.

Moore, B. C. J. (1997). *An Introduction to the Psychology of Hearing*, Academic Press, London.

Moore, B. C. J., Oldfield, S. R., and Dooley, G. J. (1989). "Detection and discrimination of spectral peaks and notches at 1 and 8 kHz," *J. Acoust. Soc. Am.,* 85:820-836.

Moore, C. N., Casseday, J. H., and Neff, W. D. (1974). "Sound localization: The role of the commissural pathways of the auditory system of the cat," *Brain Res.,* 82:13-26.

Morimoto, M. (2001). "The contribution of two ears to the perception of vertical angle in sagittal planes," *J. Acoust. Soc. Am.,* 109:1596-1603.

Morimoto, M., Iida, K., and Itoh, M. (2003). "Upper hemisphere sound localization using head-related transfer functions in the median plane and interaural differences," *Acoust. Sci and Tech.,* 24:267-275.

Müller, M. M., Malinowski, P., Gruber, T., and Hillyard, S. A. (2003). "Sustained division of the attentional spotlight," *Nature,* 424:309-312.

Nelken, I., Kim, P. J., and Young, E. D. (1997). "Linear and nonlinear spectral integration in Type IV neurons of the dorsal cochlear nucleus. II. Predicting responses with the use of nonlinear models," *J. Neurophysiol.,* 78:800-811.

Nelken, I. and Young, E. D. (1994). "Two separate inhibitory mechanisms shape the responses of dorsal cochlear nucleus type IV units to narrowband and wideband stimuli," *J. Neurophysiol.,* 71:2446-2462.

Noble, W., Byrne, D., and Ter-Host, K. (1997). "Auditory localization, detection of spatial separateness, and speech hearing in noise by hearing impaired listeners," *J. Acoust. Soc. Am.,* 102:2343-2352.

Noble, W. and Perrett, S. (2002). "Hearing speech against spatially separate competing speech versus competing noise," *Percept. Psychophys.,* 64:1325-1336.

Oliver, D. L., Beckius, G. E., Bishop, D. C., Loftus, W. C., and Batra, R. (2003). "Topography of interaural temporal disparity coding in projections of medial superior olive to inferior colliculus," *J. Neurosci.,* 23:7438-7449.

Oliver, D. L. and Huerta, M. F. (1992). "Inferior and superior colliculi," in *The Mammalian Auditory Pathway: Neuroanatomy,* ed. D. B. Webster, A. N. Popper and R. R. Fay, Springer-Verlag, New York, 168-221.

Overholt, E., Rubel, E. W., and Hyson, R. L. (1992). "A circuit for coding interaural time differences in the chick brainstem," *J. Neurosci.,* 12:1698-1708.

Palmer, A. R. and King, A. J. (1982). "The representation of auditory space in the mammalian superior colliculus," *Nature,* 299:248-249.

Palmer, A. R. and King, A. J. (1985). "A monaural space map in the guinea-pig superior colliculus," *Hearing Res.,* 17:267-280.

Perrett, S. and Noble, W. (1995). "Available response choices affect localization of sound," *Percept. Psychophys.,* 57:150-158.

Perrott, D. R. (1984). "Concurrent minimum audible angle: A re-examination of the concept of auditory spatial acuity," *J. Acoust. Soc. Am.,* 75:1201-1206.

Perrott, D. R. and Saberi, K. (1990). "Minimum audible angle thresholds for sources varying in both elevation and azimuth," *J. Acoust. Soc. Am.,* 87:1728-1731.

Pickles, J. O. (1988). *An Introduction to the Physiology of Hearing*, Academic Press, London.

Plenge, G. (1974). "On the differences between localization and lateralization," *J. Acoust. Soc. Am.,* 56:944-951.

Plenge, G. and Brunschen, G. (1971). "*A priori* knowledge of the signal when determining the direction of speech in the median plane," in *Proc. 7th Int. Congress on Acoust.*, 19, H10.

Poganiatz, I., Nelken, I., and Wagner, H. (2001). "Sound-localization experiments with barn owls in virtual space: Influence of interaural time difference on head-turning behavior," *J. Assoc. Res. Otolaryng.,* 2:1-21.

Pollack, I. and Trittipoe, W. J. (1959a). "Binaural listening and interaural noise cross correlation," *J. Acoust. Soc. Am.,* 31:1250-1252.

Pollack, I. and Trittipoe, W. J. (1959b). "Interaural noise correlations: Examination of variables," *J. Acoust. Soc. Am.,* 31:1616-1618.

Poon, P. W. and Brugge, J. F. (1993). "Sensitivity of auditory nerve fibers to spectral notches," *J. Neurophysiol.,* 20:655-665.

Pralong, D. and Carlile, S. (1994). "Measuring the human head-related transfer functions: A novel method for the construction and calibration of a miniature "in-ear" recording system," *J. Acoust. Soc. Am.,* 95:3435-3444.

Pratt, C. C. (1930). "The spatial character of high and low tones," *J. Exp. Psych.,* 13:278-285.

Rakerd, B., Hartmann, W. M., and McCaskey, T. L. (1999). "Identification and localization of sound sources in the median sagittal plane," *J. Acoust. Soc. Am.,* 106:2812-2820.

Reale, R. A. and Brugge, J. F. (2000). "Directional sensitivity of neurons in the primary auditory (A1) cortex of the cat to successive sounds ordered in time and space," *J. Neurophysiol.,* 84:435-450.

Recanzone, G. H., Makhamra, S. D. D. R., and Guard, D. C. (1998). "Comparison of relative and absolute sound localization ability in humans," *J. Acoust. Soc. Am.,* 103:1085-1097.

Rhodes, G. (1987). "Auditory attention and the representation of spatial information," *Percept. Psychophys.,* 42:1-14.

Ricard, G. L. and Meirs, S. L. (1994). "Intelligibility and localization of speech from virtual directions," *Human Factors,* 36:120-128.

Roffler, S. K. and Butler, R. A. (1967). "Factors that influence the localization of sound in the vertical plane," *J. Acoust. Soc. Am.,* 43:1255-1259.

Saberi, K., Dostal, L., Sadralodabai, T., Bull, V., and Perrott, D. R. (1991a). "Free-field release from masking," *J. Acoust. Soc. Am.,* 90:1355-1370.

Saberi, K., Dostal, L., Sadralodabai, T., and Perrott, D. R. (1991b). "Minimum audible angles for horizontal, vertical, and oblique orientations: Lateral and dorsal planes," *Acustica,* 75:57-61.

Sach, A. J., Hill, N. I., and Bailey, P. J. (2000). "Auditory spatial attention using interaural time differences," *J. Exp. Psychol. Human Percept. Perform.,* 26:717-729.

Sandel, T. T., Teas, D. C., Feddersen, W. E., and Jeffress, L. A. (1955). "Localization of sounds from single and paired sources," *J. Acoust. Soc. Am.,* 27:842-852.

Sanders, A. F. (1998). *Elements of Human Performance: Reaction Processes and Attention in Human Skill*, Lawrence Erlbaum Associates, New Jersey.

Schnupp, J. W. H. and King, A. J. (1997). "Coding for auditory space in the nucleus of the brachium of the inferior colliculus in the ferret," *J. Neurophysiol.,* 78:2717 - 2731.

Shaw, E. A. G. (1974). "The external ear," in *Handbook of Sensory Physiology,* ed. W. D. Keidel and W. D. Neff, Springer-Verlag, New York, 455-490.

Shinn-Cunningham, B., Ihlefeld, A., and Schoolmaster, M. (2004). "Selective and divided attention: Extracting information from simultaneous sound sources," in *Proc. Int. Conf. Auditory Display*, Sydney, Australia, submitted.

Shinn-Cunningham, B. and Kawakyu, K. (2003). "Neural representation of source direction in reverberant space," in *Proc. IEEE Workshop on Applications of Signal Processing to Audio and Acoustics*, New Paltz, USA, 79-82.

Shinn-Cunningham, B. G. (2000). "Learning reverberation: Considerations for spatial auditory displays," in *Proc. Int. Conf. Auditory Display*, Atlanta, USA, 126-134.

Shinn-Cunningham, B. G., Santarelli, S., and Kopco, N. (2000). "Tori of confusion: Binaural localization cues for sources within reach of a listener," *J. Acoust. Soc. Am.,* 107:1627-1636.

Slattery, W. H. and Middlebrooks, J. C. (1994). "Monaural sound localization: Acute verses chronic unilateral impairment," *Hearing Res.,* 75:38-46.

Smith, P. H., Joris, P. X., and Yin, T. C. (1993). "Projections of physiologically characterized spherical bushy cell axons from the cochlear nucleus of the cat: Evidence for delay lines to the medial superior olive," *J. Comp. Neurol.,* 331:245-260.

Spangler, K. M. and Warr, W. B. (1991). "The descending auditory system," in *Neurobiology of Hearing: The Central Auditory System,* ed. R. A. Altschuler, R. P. Bobbin, B. M. Clopton and D. W. Hoffman, Raven Press, New York, 27-45.

Spence, C. J. and Driver, J. (1994). "Covert spatial orienting in audition: Exogenous and endogenous mechanisms," *J. Exp. Psych. Human Percept. Perform.,* 20:555-574.

Spieth, W., Curtis, J. F., and Webster, J. C. (1954). "Responding to one of two simultaneous messages," *J. Acoust. Soc. Am.,* 26:391-396.

Spitzer, M. W. and Semple, M. N. (1995). "Neurons sensitive to interaural phase disparity in gerbil superior olive: Diverse monaural and temporal response properties," *J. Neurophysiol.,* 73:1668-1690.

Stelmachowicz, P. G., Pittman, A. L., Hoover, B. M., and Lewis, D. E. (2001). "Effect of stimulus bandwidth on the perception of /s/ in normal and hearing-impaired children and adults," *J. Acoust. Soc. Am.,* 110:2183-2190.

Sterbing, S. J., Hartung, K., and Hoffmann, K. P. (2003). "Spatial tuning to virtual sounds in the inferior colliculus of the guinea pig," *J. Neurophysiol.,* 90:2648-2659.

Stevens, S. S. and Newman, E. B. (1936). "The localization of actual sources of sound," *Am. J. Psychol.,* 48:297-306.

Strybel, T. Z. and Fujimoto, K. (2000). "Minimum audible angles in the horizontal and vertical planes: Effects of stimulus onset asynchrony and burst duration," *J. Acoust. Soc. Am.,* 108:3092-3095.

Suga, N. and Ma, X. (2003). "Multiparametric corticofugal modulation and plasticity in the auditory system," *Nat. Reviews Neurosci.,* 4:783-794.

Suzuki, Y. and Sone, T. (1986). "Influence of an interfering noise on the localization of a band noise source," in *Proc. 12th Int. Congress on Acoust.*, Toronto, Canada, B2-10.

Suzuki, Y., Yokoyama, T., and Sone, T. (1993). "Influence of interfering noise on the sound localization of a pure tone," *J. Acoust. Soc. Jpn.,* 14:327-339.

Takahashi, T. T. and Keller, C. H. (1994). "Representation of multiple sound sources in the owls auditory space map," *J. Neurosci.,* 14:4780-4793.

Teder, W. and Naatanen, R. (1994). "Event-related potentials demonstrate a narrow focus of auditory spatial attention," *Neuroreport,* 5:709-711.

Teder-Sälejärvi, W. A., Hillyard, S. A., Röder, B., and Neville, H. J. (1999). "Spatial attention to central and peripheral auditory stimuli as indexed by event-related potentials," *Cog. Brain Res.,* 8:213-227.

Thiran, A. B. and Clarke, S. (2003). "Preserved use of spatial cues for sound segregation in a case of spatial deafness," *Neuropsychologia,* 41:1254-1261.

Thurlow, W. R., Marten, A. E., and Bhatt, B. J. (1965). "Localization aftereffects with pulse-tone and pulse-pulse stimuli," *J. Acoust. Soc. Am.,* 37:837-842.

Tobias, J. V. and Schubert, E. D. (1959). "Effective onset duration of auditory stimuli," *J. Acoust. Soc. Am.,* 31:1595-1605.

Tollin, D. J. (2003). "The lateral superior olive: A functional role in sound source localization," *Neuroscientist,* 9:127-143.

Trahiotis, C. and Bernstein, L. R. (1990). "Detectability of interaural delays over select spectral regions: Effects of flanking noise," *J. Acoust. Soc. Am.,* 87:810-813.

van Schaik, A., Jin, C., and Carlile, S. (1999). "Human sound localization of band-pass filtered noise," *Int. J. Neural Systems,* 9:441-446.

Vickers, D. A., Moore, B. C. J., and Baer, T. (2001). "Effects of low-pass filtering on the intelligibility of speech in quiet for people with and without dead regions," *J. Acoust. Soc. Am.,* 110:1164-1175.

von Békésy, G. (1960). *Experiments in Hearing*, McGraw-Hill, New York.

Wallach, H. (1940). "The role of head movements and vestibular and visual cues in sound localization," *J. Exp. Psychol.,* 27:339-368.

Wenzel, E. M., Arruda, M., Kistler, D. J., and Wightman, F. L. (1993). "Localization using non-individualized head-related transfer functions," *J. Acoust. Soc. Am.,* 94:111-123.

Whitehead, M. L., Simons, I., Stagner, B. B., and Martin, G. K. (1997). "The frequency response of the ER-2 speaker at the eardrum," *J. Acoust. Soc. Am.,* 101:1195-1198.

Wightman, F. L. and Kistler, D. J. (1989a). "Headphone simulation of free field listening II: Psychophysical validation," *J. Acoust. Soc. Am.,* 85:868-878.

Wightman, F. L. and Kistler, D. J. (1989b). "Headphone simulation of free-field listening I: Stimulus synthesis," *J. Acoust. Soc. Am.,* 85:858-867.

Wightman, F. L. and Kistler, D. J. (1992). "The dominant role of low-frequency interaural time differences in sound localization," *J. Acoust. Soc. Am.,* 91:1648-1661.

Wightman, F. L. and Kistler, D. J. (1997). "Monaural sound localization revisited," *J. Acoust. Soc. Am.,* 101:1050-1063.

Wightman, F. L. and Kistler, D. J. (1999). "Resolution of front-back ambiguity in spatial hearing by listener and source movement," *J. Acoust. Soc. Am.,* 105:2841-2853.

Wise, L. Z. and Irvine, D. R. F. (1983). "Auditory response properties of neurons in deep layers of cat superior colliculus," *J. Neurophysiol.,* 49:674-685.

Wise, L. Z. and Irvine, D. R. F. (1985). "Topographic organization of interaural intensity difference sensitivity in deep layers of cat superior colliculus: Implications for auditory spatial representation," *J. Neurophysiol.,* 54:185-211.

Woldorff, M. G. and Hillyard, S. A. (1991). "Modulation of early auditory processing during selective listening to rapidly presented tones," *Electroenceph. and Clin. Neurophysiology,* 79:170-191.

Woods, W. S. and Colburn, H. S. (1992). "Test of a model of auditory object formation using intensity and interaural time difference discrimination," *J. Acoust. Soc. Am.,* 91:2894-2902.

Woodworth, R. S. (1938). *Experimental Psychology*, Holt, Rinehart and Winston, New York.

Woodworth, R. S. and Schlosberg, H. (1954). *Experimental Psychology*, Methuen and Co. Ltd., London.

Worley, J. W. and Darwin, C. J. (2002). "Auditory attention based on differences in median vertical plane position," in *Proc. Int. Conf. Auditory Display*, Kyoto, Japan, 440-444.

Wurtz, R. H. and Albano, J. E. (1980). "Visual-motor function of the primate superior colliculus," *Annu. Rev. Neurosci.,* 3:189-226.

Xu, L., Furukawa, S., and Middlebrooks, J. C. (1998). "Sensitivity to sound-source elevation in nontonotopic auditory cortex," *J. Neurophysiol.,* 80:882-894.

Yin, T. C. and Chan, J. C. (1990). "Interaural time sensitivity in the medial superior olive of cat," *J. Neurophysiol.,* 64:465-488.

Yin, T. C. T., Kuwada, S., and Sujaku, Y. (1984). "Interaural time sensitivity of high-frequency neurons in the inferior colliculus," *J. Acoust. Soc. Am.,* 76:1401-1410.

Yost, W. A. (1976). "Lateralization of repeated filtered transients," *J. Acoust. Soc. Am.,* 60:178-181.

Yost, W. A., Dye Jr., R. H., and Sheft, S. (1996). "A simulated "cocktail party" with up to three sound sources," *Percept. Psychophys.,* 58:1026-1036.

Young, E. D., Spirou, G. A., Rice, J. J., and Voigt, H. F. (1992). "Neural organization and responses to complex stimuli in the dorsal cochlear nucleus," *Phil. Trans. Royal Soc. London B,* 336:407-413.

Zahorik, P. (2000). "Limitations in using Golay codes for head-related transfer function measurement," *J. Acoust. Soc. Am.,* 107:1793-1796.

Zahorik, P. (2002). "Assessing auditory distance perception using virtual acoustics," *J. Acoust. Soc. Am.,* 111:1832-1846.

Zahorik, P. and Wightman, F. L. (2001). "Loudness constancy with varying sound source distance," *Nat. Neurosci.,* 4:78-83.

Zattore, R. J., Bouffard, M., Ahad, P., and Belin, P. (2002). "Where is 'where' in the human auditory cortex?," *Nature Neurosci.,* 5:905-909.

Zhou, B., Green, D. M., and Middlebrooks, J. C. (1992). "Characterization of external ear impulse responses using Golay codes," *J. Acoust. Soc. Am.,* 92:1169-1171.

Zurek, P. M. (1985). "Spectral dominance in sensitivity to interaural delay for broadband stimuli," *J. Acoust. Soc. Am.,* 78:S18.

Zurek, P. M. (1987). "The precedence effect," in *Directional Hearing,* ed. W. A. Yost and G. Gourevitch, Springer-Verlag, New York, 85-105.

Zurek, P. M. (1990). "Binaural advantages and directional effects in speech intelligibility," in *Acoustical Factors Affecting Hearing Aid Performance,* ed. G. A. Studebaker and I. Hochberg, Allyn and Bacon, Boston, 255-276.