# Chapter one: Introduction

Human beings, like other organisms, obtain visual and auditory information to help them know about the nature and whereabouts of objects and events within their environment. Auditory information has several advantages over visual information in that it is available to a person no matter what direction the person is facing, what the visibility conditions are, or whether opaque objects lie between the person and the position of sound production. For this reason auditory information has the potential to inform us about the nature and whereabouts of objects and events we could not otherwise obtain.

Auditory localization has been studied in the laboratory in various ways, and has been shown to involve the pick-up of various forms of acoustical information including interaural time and level differences, and pinna-induced spectral cues. Typically, auditory localization research involves experiments in which listeners are presented sounds from variously positioned loudspeakers and are required to judge the direction of the sounds. The level of accuracy under various conditions is assessed. This form of experimentation has the greatest potential to represent localization performance in everyday conditions. Another approach has been to test listeners in their ability to detect relative differences in the angular position of two sound source positions. This method is referred to as minimal audible angle (MAA) research and directly assesses the spatial acuity of the auditory system.

A further way of investigating auditory localization ability has involved a method called lateralization. This approach involves presenting sounds over headphones, which allows precise control and investigation of interaural differences. However, not all parameters involved in auditory localization are represented in lateralization experiments (Wightman & Kistler, 1993). Because pinna induced spectral cues are ignored, the sound image is experienced as occurring inside the head. Listeners are required to make judgements about the relative position of this intra-cranial sound image. Therefore the findings of such research are not highly representative of auditory localization occurring in the everyday world.

Finally, a form of auditory localization research that has recently been developed involves the use of virtual sound sources, presented to listeners over headphones. This form of headphone presentation differs greatly from that used in lateralization studies in that spectral information is also represented. Listeners experience virtual sound sources as occurring naturally outside the head. Such research has the potential to allow precise control over all potential cues to sound localization. However, it is still the case that listeners localize sounds more accurately under normal (non-virtual) conditions, therefore some cues to sound source position may not be represented accurately even in virtual acoustical environments.

Most research into auditory localization is conducted on listeners who are required to remain absolutely motionless while they listen to the sound. The outcome of such research suggests that acoustic information concerning the whereabouts of sound sources may often be ambiguous, especially if the sound has limited frequency content, as is often the case in everyday situations. Wallach (1939; 1940) developed a theory that suggested changes in acoustic information occurring coincidentally with head movements could disambiguate otherwise ambiguous localization cues.

It is reasonable to assume that under normal conditions, movements of a listener's head are usual when sound sources are being localized. For example, if a person hears someone whistling, and wants to see the whereabouts of the whistler, they will probably need to move their head to bring the whistler into their line of sight. Heffner and Heffner (1992) have provided evidence that suggests a major function of auditory localization is to allow the field of best vision to be directed towards the source of a sound. They showed that across a broad range of mammal species, including humans, horizontal width of field of best vision is highly correlated with sound localization acuity; the narrower the field of best vision, the lower the sound localization threshold (r=.9). Compared with other mammals, humans have a very narrow field of best vision and very acute sound localization.

Perrott, Saberi, Brown and Strybel (1990) showed that a primary role of the spatial sensitivity of the human auditory system may indeed be to enable listeners to regulate their visual gaze, showing that search time for a visual target is significantly reduced if a sound emanates from the position of the target. They also provided evidence that head movements are a significant factor in aurally guided visual search, even for events within 20° of a person's initial line of sight. Such movements will usually bring about changes in the acoustic energy picked up by the listener.

In a series of ingenious simulation experiments, Wallach (1939; 1940) was able to provide some convincing evidence in support of his head motion localization theory. However, there is little direct evidence that head movements actually assist human auditory localization. One reason for this may be that essential elements of Wallach's theory, as well as localization theory in general, have been overlooked in the design of most subsequent experiments that have investigated the role of head movement in relation to this sensory function.

## *1.1. STATIC LOCALIZATION CUES*

### 1.1.1. Classical interaural cues

According to classical theory, the binaural form of the human auditory system is the basic anatomical feature underlying horizontal (left-right) localization. The arrangement of the ears, on opposite sides of the head, means the acoustic energy emanating from a sound source is accessible from two separate points in space. The relationships between characteristics of the acoustic energy at one ear relative to the other provide the listener with what are termed interaural cues, and these are assumed to enable horizontal (left-right) localization. Classical theory suggests that low-frequency acoustic energy provides interaural time/phase cues and high-frequency acoustic energy provides interaural sound pressure level cues.

Interaural relationships in time vary as a function of the relative distance of the sound source from each of the ears. If a sound occurs directly in front of a listener (in what is termed the median vertical plane or MVP — see Figure 1, p.

8), the distance from the sound source to each of the ears is exactly the same, so the sound will arrive at both ears at the same time, with the phase of the signal being identical at both ears. However, if the sound occurs, for example, somewhere to the left, the distance from the source to the left ear is somewhat less than to the right ear, so the sound will arrive at the left ear slightly before it arrives at the right, and the phase of the signal will be more advanced at the left ear than at the right. The interaural time difference (ITD) increases the more the sound source is displaced from the MVP. The maximum ITD, occurring for a sound directly left or directly right of the listener, is approximately 700-μs (Middlebrooks and Green, 1991). The auditory system has been shown to be extremely sensitive to ITDs, with listeners able to detect ITDs of approximately 10-μs (Klumpp and Eady, 1956).

Interaural relationships in sound pressure level depend on what has been called the head's sound shadow effect, which involves the head acting as an acoustic barrier. If a sound occurs from directly in front of a listener, the action of the head, as an acoustic barrier is relatively minor and essentially identical at each ear, so the sound pressure level is the same at each ear. If a sound occurs somewhat to the left or to the right, the direct path of the sound to the far ear is blocked by the head, causing a reduction in the level of the sound at the far ear. The greater the angular displacement of the source from a listener's MVP the greater the head's sound shadow effect. Listeners can detect interaural sound pressure level differences (ILDs) of as little as 0.5 dB (Mills, 1960). When the source is directly left or directly right of the listener, interaural level differences can be as much as 20 dB for 4-kHz tones and 35 dB for 10-kHz tones (Middlebrooks Makous & Green, 1989). At frequencies lower than about 1 kHz, ILDs become negligible (Middlebrooks & Green, 1991). This is because wavelengths at lower frequencies are sufficiently long to allow the wavefront to diffract (bend) around an object the size of the human head.

## 1.1.2. The ambiguity of classical interaural cues

The classical model assumes the head to be a perfect sphere and always remaining static. The ears are considered as merely holes centred on opposite sides of the spherical head. Using this model, interaural cues specify the angle of horizontal (left-right) displacement of a sound source from the median vertical plane (MVP). The vertex of this angle lies at the centre of the listener's head. As such, interaural cues usually do not limit the potential position of the source to one particular direction, but specify a range of directions. For example, interaural cues might indicate the horizontal displacement of the source is 60° to the left of the MVP. As shown in Figure 1, such an angle encompasses a cone-shaped locus covering positions forward, rearward, above and below the interaural axis. This and other cone-shaped loci, of various angle size, have been termed *cones of*

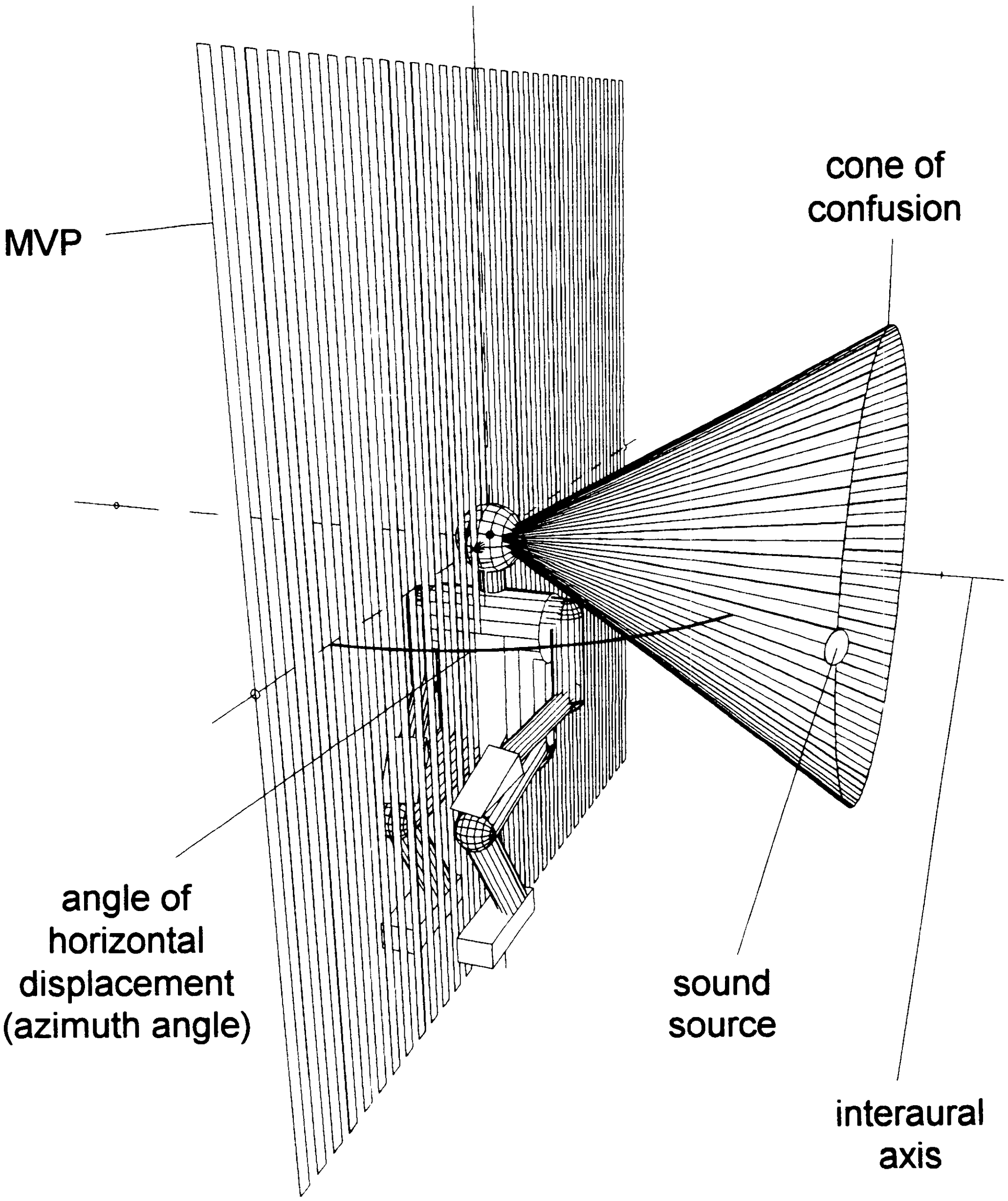

Figure 1. A horizontal angle of displacement of 60° to the left of the median vertical plane, specifying a cone shaped locus (*cone of confusion*) encompassing forward, rearward, upward and downward directions.

*confusion* (Woodworth & Schlosberg, 1954). Thus, considered purely in terms of the geometry of the classical model, interaural cues alone do not specify the elevation of the source, nor whether it is forward or rearward.

A classical model does not take into account that the human head is not actually a perfect sphere, and human ears are not just holes and are not centred on exactly opposite sides. Nevertheless, measurements made within the ear canals of actual listeners (Wightman and Kistler, 1993; 1994), show that the interaural phase and level information is essentially the same as that predicted by the classical model. Figures 2 and 3 show actual interaural information measured for a typical listener (Wightman and Kistler, 1994) and plotted as contours of constant interaural difference. Although not perfectly centred around the interaural axis and somewhat irregular in shape, the degree to which these contours correspond with those of the classical model, suggest the classical model's validity. Therefore, if classical interaural cues were the only source of information, front-back, up-down localization would be impossible.

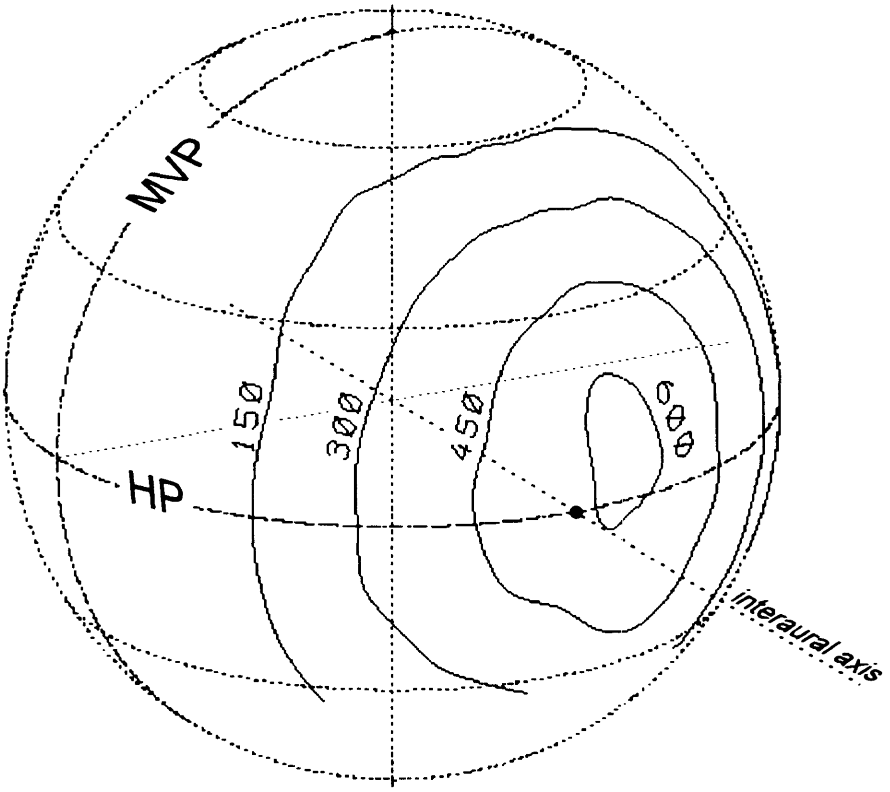

Figure 2. Contours of constant interaural time difference (ITD, in microseconds) extracted from head related transfer function (HRTF) measurements by estimating the delay at the maximum in the cross-correlation between left and right ear HRTFs (adapted from Wightman and Kistler, 1994).

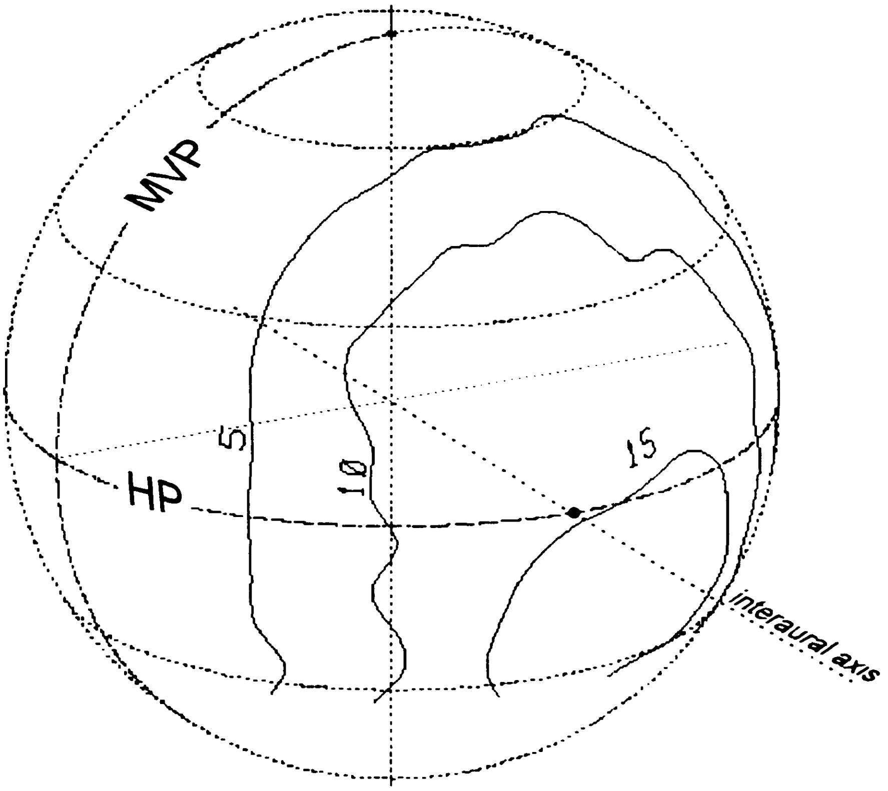

Figure 3. Contours of constant interaural level difference (ILD, in dB) obtained by subtracting the overall level of the HRTF (200 Hz - 14 kHz, in dB) in one ear from the overall level in the other ear (adapted from Wightman and Kistler, 1994).

### 1.1.3. Representing auditory space with a convenient co-ordinate system

Under the classical model, the horizontal angle of displacement specified by classically defined interaural cues has properties consistent with the azimuth angle defined in the double-pole co-ordinate system used by Middlebrooks Makous and Green (1989). Because of this consistency, a double-pole co-ordinate system is convenient for referring to the spatial geometry assumed by a classical model. A double-pole system is used throughout this report and is illustrated schematically in Figure 4. Elevation in this system refers to the angle formed between the HP and a line projected from the centre of the listener's head to a given point in space. Front and back positions comprise identical azimuth angles, therefore co-ordinates are always referred to as lying in a particular hemisphere (front or back).

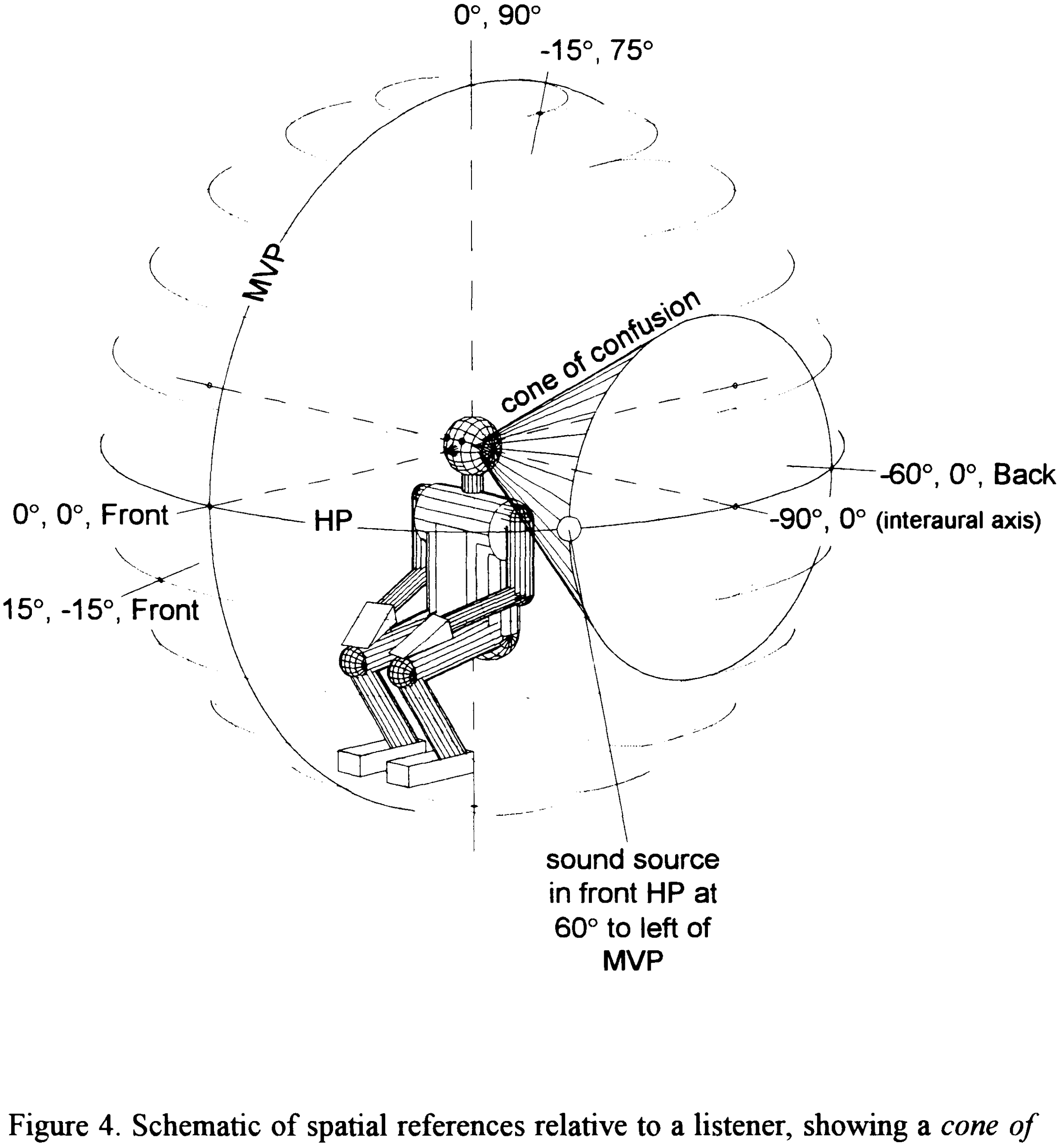

Figure 4. Schematic of spatial references relative to a listener, showing a *cone of confusion* at a leftward azimuth angle of 60°. Increments by 15° in azimuth angle are represented on a sphere as circles perpendicular to the horizontal plane (HP). Azimuth angles are positive when rightward, negative when leftward. Increments by 15° in elevation angle are shown as circles perpendicular to the median vertical (MVP). These are positive when upward, negative when downward. Examples are shown of co-ordinates using this *double-pole* system.

## 1.1.4. High-frequency pinna derived spectral cues

As early as the late nineteenth century, Mach (1874, cited in Thompson, 1882) speculated that the convolutions of the pinnae play a role in auditory localization. Speculation continued until the 1960s when Batteau (1968) reported that listeners could accurately judge elevation and make correct front-back discriminations while listening via headphones to sounds presented in a separate room. The sounds were picked up binaurally by two extremely high fidelity microphones fitted with casts of real pinnae. If the pinnae were removed localization accuracy broke down.

Fisher and Freedman (1968) devised a follow-up experiment in which listeners were asked to localize pulsed white noise presented from sources in the HP, forward and rearward of the interaural axis, with and without tubes placed in their ears. The assumption was that the tubes would disrupt any effects the pinnae might have. The results showed superior front-back localization without the tubes than with them. The importance of the pinnae was then clearly demonstrated when casts of real pinnae were placed at the ends of the tubes. Under this condition, front-back localization was restored to about the same level of accuracy as without the tubes.

Shaw and Teranishi (1968) showed that the pinnae act as directionally dependent filters, transforming the spectra of sounds differently depending on the direction of the incident sound. They observed that the pinnae created a series of direction dependent peaks and notches in the spectra of signals from about 4 kHz upward. A subsequent profusion of behavioural research has shown that this spectral information enables motionless listeners to unambiguously localize sounds that include complex high-frequency energy (e.g., Blauert, 1969/70; Butler & Helwig, 1983; Hebrank & Wright, 1974)

For the auditory system to make use of pinna cues, the sound must comprise energy spanning a range of frequencies within the 4-12-kHz region (Hebrank & Wright, 1974). When acoustic energy above 4 kHz is absent, as in noise that is low-pass filtered at 2 kHz, then localization errors occur of a sort expected from cone-of-confusion principles (Perrett & Noble, 1995).

The usefulness of spectrally derived information may also depend on the smoothness of a sound's initial spectrum, or prior knowledge of the sound's initial spectrum. Wightman and Kistler (1997) argue that sounds occurring in everyday listening conditions are of uncertain spectral composition and that this may interfere with spectral cue extraction. When Hebrank and Wright (1974) introduced randomly placed peaks and notches into the spectrum of broadband noise so that the spectrum of the sound was different from one trial to the next, they found localization of MVP sources was significantly disrupted, compared to when sounds with constantly flat spectra were involved. Wightman and Kistler (1997) employed a technique of spectral scrambling which randomly varied the level of the signal in each critical band within a range of 40 dB. Sounds were presented from free-field sources positioned all around the listener from -48° to +72° elevation. Compared to localization of sounds with a constantly flat spectrum, listeners' performance was typically degraded in front-back and elevation dimensions.

As might be expected, interaural information was barely disrupted by the spectral scrambling, as indicated by highly proficient performance in the azimuth dimension — apart from some minor degradation in azimuth for sources near the theoretical interaural axis, probably because actual cones of confusion are not centred exactly around this axis. If spectral uncertainty is a problem for localization of sounds in everyday conditions, then spectral cues may be of limited use in disambiguating interaural information.

Gilkey and Anderson (1995) compared speech (266 different words — spoken by a male and female) with click stimuli in relation to motionless listeners' abilities to localize sounds, and found cone-of-confusion type errors were more pronounced with the speech stimuli. It is not clear whether the increased error was due to reduced high-frequency content, spectral uncertainty, or a further possibility, raised by Gilkey and Anderson (1995), namely, ongoing changes in the high-frequency spectra. Whatever the reason, the results provide further evidence that pinna derived spectral cues may be less than optimal for localization of types of sound we encounter in everyday life.

## 1.1.5. Low-frequency shoulder/torso-derived spectral cues

There is evidence that spectral information for source whereabouts may be derivable from low-frequency acoustic energy. Gardner (1973) carried out experiments in which the function of the pinnae was disrupted. This was achieved by occluding pinnae cavities while allowing a free passageway to the ear canal. He found that with broadband and 1-kHz high-pass noises, listeners achieved fairly accurate localization of sources arrayed from -18° to +18° elevation in the front MVP. Accuracy was noticeably degraded with a 2-kHz high-pass noise and even more degraded with a 3-kHz high-pass noise. Gardner measured acoustic spectra from inside the ears of a mannequin (head and torso). When he compared measurements associated with a source positioned at +18° and a source positioned at -18° elevation, he found substantial spectral differences within the range of 0.7 to 3 kHz. These differences disappeared when the torso of the mannequin was removed, suggesting the torso acts as a directionally dependent filter for acoustic energy at frequencies not affected by the pinnae.

Kuhn (1987) carried out extensive acoustical measurements using a KEMAR (Knowles Electronics Manikin for Acoustical Research) and found directionally dependent spectral differences in the spectral region below 2 kHz, occurring for source positions in the MVP. Asano, Suzuki and Sone (1990) filtered broadband noise in accordance with measured head related transfer functions (HRTF) to successfully create, via headphones, virtual source positions in the upper MVP. It was found that if small peaks and dips in the HRTFs were smoothed in the frequency range above 2 kHz, localization of the resulting virtual sources was about as accurate as when HRTFs were not smoothed. Smoothing of the HRTFs in the frequency region below 2 kHz, by contrast, resulted in a large increase in the number of front-back confusions.

There is also evidence that low-frequency energy enables a degree of front-back discrimination within the HP. With a 4-kHz low-pass noise presented from HP sources, Musicant and Butler (1984) observed front-back confusions on just 8.1% of localization trials involving motionless listeners. With the same signal, occlusion of the pinnae cavities had virtually no effect on front-back discrimination, with 8.8% of trials producing front-back confusion. With a 1-kHz low-pass noise the front-back confusion rate was much greater — 36.8% with unoccluded pinnae and 41.4% with occluded pinnae. These results strongly suggest that information for front-back discrimination can be obtained from acoustic energy within the range of 1 - 4 kHz, and that such information is not associated with the pinnae.

Evidence provided by Musicant and Butler suggests that low-frequency spectral cues, are not as salient and do not carry as much weight as high-frequency spectral cues. With pinnae unoccluded, front-back confusion occurred on only 1.5% of 4-kHz high-pass noise trials, a substantial reduction compared with the 8.1% occurring with 4-kHz low-pass noise. With the pinnae occluded, front-back discrimination with 4-kHz high-pass noise occurred on 39.5% of trials, indicating the strong involvement of the pinnae in front-back discrimination with such a signal. With a broadband noise, front-back confusion occurred on 0.5% of trials with unoccluded pinnae and 23.9% of trials with occluded pinnae. Since front-back confusion occurred on only 8.8% of trials with the 4-kHz low-pass noise when pinnae were occluded, it seems that high-frequency spectral information may outweigh low-frequency spectral information.

## *1.2. HEAD MOVEMENT CUES*

### 1.2.1. Theory

Several researchers have hypothesized that head movements assist localization. Van Soest (1929, cited in Blauert, 1983) recognized that if a listener rotates the head about a vertical axis, interaural information is transformed. He pointed out that with the same head rotation, the pattern of changing interaural time differences occurring for a forward source, is reversed for a rearward source, thus providing a cue for front-back discrimination.

Wallach (1939; 1940) formulated a more extensive theory of sound localization based on head motion. He proposed that changes in classical interaural information occurring with head motion could be used to provide completely unambiguous cues to the whereabouts of sound sources. Like van Soest, Wallach suggested that rotating the interaural axis through the horizontal plane would eliminate front-back ambiguity due to the contrasting alteration in interaural information for sources behind versus those in front of that axis. Moreover, Wallach also argued that the same kind of rotation would provide cues to source elevation. When a sound source is located in the HP, a given head rotation brings about the same change in azimuth angle; for example, 15° of head rotation will bring about a 15° change in azimuth angle. On the other hand, rotation produces no change in azimuth angle when a sound is directly overhead (or directly below) — that angle remains 0° for any amount of head rotation. For intermediate elevations, the change in azimuth angle is somewhat greater than zero but somewhat less than the degree of head rotation. Wallach proposed that the rate of change in azimuth angle relative to the rate of change in head orientation could allow listeners to judge the vertical displacement of the sound source above or below the HP. Furthermore, the remaining ambiguity between above and below could be eliminated by tilting the head from side to side.

### 1.2.2. Wallach's empirical evidence

Wallach (1939; 1940) artificially created changing interaural information using an apparatus involving an array of 20 loudspeakers and a rotary switch attached, by its shaft, to a listener's rotating head. The signal (piano or orchestral music from victrola records) could be directed, via the contacts of the rotary switch, from one loudspeaker to the next. In one experiment (Wallach, 1940) loudspeakers were arranged in the front HP and the shaft of the rotary switch was attached to the top of the listener's head. While the signal occurred, the listener (maintaining a fixed posture) sat passively on a rotatable seat which was repeatedly rotated, one way then the other, by the experimenter. Using this procedure, the apparatus artificially produced the changing azimuth angles that would normally arise for a rotating listener if the sound was generated from directly behind. All of the five participating listeners heard the sound as coming from behind rather than from the loudspeaker array to the front. In another part of the experiment, the rotary switch was by-passed with the signal always emanating from a single loudspeaker in front of the listener. This time the passively rotated listeners heard the sound as coming from the direction of the actual source. Such an outcome clearly suggests that a front-back cue is derivable from the changing interaural information arising from head rotation.

Using the same apparatus, Wallach (1939; 1940) also simulated various source elevations. By varying the separation between the loudspeakers arrayed in the HP, head rotation could be made to generate various rates of change in azimuth angle. In one experiment (Wallach, 1939) the array was configured so that the signal was switched to whatever loudspeaker coincided with the direction in which the listener was facing at any given moment. In this way the azimuth angle specified by the interaural information remained close to 0° throughout head rotation. Ten listeners participated in the experiment after being selected from a larger group of 17, on the basis of their having the ability to localize sounds produced from actual sources positioned overhead. While the signal occurred, the listeners actively made head rotations, and although the loudspeakers were actually in the HP, all listeners heard the sound as coming from directly above. In another experiment, the apparatus was configured to produce the rate of change in interaural information that would normally arise for a front-on source elevated 60° from the HP. That is, for every 3° of head rotation the change in azimuth angle was 1.5°. Listeners reported the sounds as coming from elevations of between 64° and 43°. Source elevations of 78° in front and 60° behind were also simulated with all listeners estimating an apparent source elevation within 10° of the theoretical direction of the simulated source. Wallach pointed out that the changes in interaural information did not rule out symmetrical positions below the HP, yet it was observed that on all but one occasion listeners perceived the sound as coming from above the HP, suggesting a strong upward bias. The evidence Wallach provided clearly implicated head motion as a potential way of obtaining cues to source elevation.

According to Wallach, the changing interaural information relative to a horizontally rotating head is such that the rate of change in azimuth angle relative to the rate of change in head rotation varies only slightly with varying elevation within 30° of the HP. He suggested that these lower elevations could not be successfully produced with his simulation apparatus, if the loudspeaker array was in the HP and rotation was about a central vertical axis. Wallach suggested that natural head movements are not limited to purely rotational ones, and that side to side tilting of the head about a central front-to-back horizontal axis would be required for judgement of sources at lower elevations.

To simulate lower elevations, the whole of the apparatus was shifted through 90° so that the 20 loudspeakers were positioned in an arc extending from left to right above the listener's head. The rotary switch was repositioned behind the head with its shaft attached to the back of the listener's head. On one occasion, listeners were tested with the loudspeakers spaced so that changes in azimuth angle, coincidental with side to side head tilting, would be consistent with those theoretically arising for a source 30° above the HP. On another occasion, the loudspeaker spacing theoretically simulated an elevation of 20°. The listeners reported hearing the sounds within about 16° of these theoretical elevations, sometimes to the front and sometimes to the back. This result suggested that elevation judgement involving sources near the HP would require the head to be turned about a horizontal (front-to-back) axis.

The evidence from these simulations suggested to Wallach that head motion cues were the principal means by which unambiguous sound localization could be achieved. He acknowledged that motionless listeners could quite reliably discriminate between front and back and credited this to what he termed the pinna factor. He suggested that because the artificially produced directions were heard in preference to the actual source directions, the pinna factor must have a subordinate role to head motion cues. At the time of Wallach's writing, little was known about the pinna cue. Wallach's conclusion may therefore not have been entirely legitimate. Current understanding would suggest that the piano or orchestral music used for the signal in the simulation experiments may not have been sufficiently rich in complex high-frequency acoustic energy to permit the generation of strong pinna cues, although it is also plausible that the equipment Wallach used may have introduced enough noise (from the recording and playback of the music) and switching transients (from activation of successive loudspeakers) to allow salient pinna cue generation. The important point, though, is that the apparent directions Wallach simulated were entirely dependent on transformations of classically defined interaural information arising coincidentally with particular types and rates of head displacement.

### 1.2.3. Non-auditory information

Wallach (1940) went on to investigate the non-auditory information that would have to be involved in the head motion localization cue. His theory implies that as well as the ongoing change in interaural information, the listener requires information about the ongoing change in head orientation in order for the head motion localization cue to function. Wallach (1940) investigated the roles of information derived from the visual and vestibular systems. Using the same sound direction simulation apparatus as before, he artificially produced the interaural cues for an apparent source at 90° elevation while listeners were passively rotated by the experimenter. Eight listeners, who could accurately localize overhead sound sources, were tested blindfolded and then not blindfolded. Wallach argued that the blindfolded condition would have provided the listener with virtually no sensory information about movement other than that provided by the vestibular system, whereas in the non-blindfolded condition, information from both the vestibular and visual systems would have been available. In the blindfolded condition, five of the eight listeners tended to localize the simulated sound source position between 10° and 20° rearward of directly above, while the others localized it directly above. In the non-blindfolded condition, all listeners localized the simulated sound source position directly overhead. Wallach took this outcome to suggest that some listeners tend to underestimate the extent of their rotation if they have to rely purely upon vestibular information, whereas when visual information is available listeners can accurately gauge the extent of their rotation.

To investigate whether visual cues alone would provide the information about head movement, Wallach used a cylindrical revolving cloth screen to create, for a motionless listener seated within, the illusion of self-rotation. Research carried out more recently has demonstrated that vision acts to provide a powerful sense of self-movement. Indeed visually derived information concerning body movement usually dominates kinaesthetic cues (Lishman & Lee, 1973). In Wallach's experiment, listeners seated inside the revolving screen experienced themselves as rotating and the screen as being at rest. Sounds were presented from outside the revolving screen, via a loudspeaker positioned directly to the front of the listener. All twelve listeners (who could accurately localize a sound presented from overhead in normal circumstances) heard the sound as being presented from directly above, as soon as, or shortly after, it had begun. Wallach concluded that visual cues alone are sufficient to provide a listener with the information about head motion that would be required for the head motion localization cue to function. This outcome suggests that localization involving head motion should be optimal under conditions in which the listener is able to see their surroundings.

## 1.2.4. Movement of a sound source about a motionless listener

Theoretically, detection of sound source movement by a motionless observer involves the same changing interaural relationships occurring for a moving observer and a stationary source. Research findings relating to the ability of a motionless listener to perceive source movement should therefore be highly relevant to the study of sound localization that is based on changing interaural relationships brought about by head movement.

Most of the research into the ability of motionless listeners to perceive the motion of sound sources has focused on measurements of the minimum audible movement angle (MAMA). The MAMA can be defined as the smallest amount a sound source must move before a motionless listener can detect the movement or detect the direction of the movement. Minimum audible movement angle research has shown that MAMAs are: 1) smallest with movement directly in front of the listener and become larger when movement occurs in regions displaced from the MVP (Chandler & Grantham, 1992); 2) with pure tone stimuli, MAMAs are smallest at lower frequencies (Perrott & Tucker, 1988); 3) Broadband signals produce the smallest MAMAs (Chandler & Grantham, 1992); and 4) MAMAs are generally smaller the lower the velocity at which the sound source moves (Perrott & Tucker, 1988).

Detection of whether a source is moving or not may relate to the ability to judge the direction of sources elevated 90° from the HP, since, according to Wallach's theory, detection of no change in interaural relationships coinciding with head rotation, specifies that the source is either directly overhead or directly below;

any change in interaural relationships means the source is at an intermediate elevation.

Detection of the direction of movement would conceivably relate to front-back discrimination since the direction of change in azimuth angles, specified by changing interaural relationships coinciding with head rotation, specifies whether a source is to the front or to the back.

The ability of listeners to discriminate between source movement velocities has also been investigated. Perrott, Buck, Waugh and Strybel (1979) have estimated that listeners can detect differences in source velocity of 5.3°/s. Velocity discrimination of source movement would seem to be relevant to discrimination between various rates of change in interaural relationships that is required for elevation judgement, according to Wallach's theory. As mentioned earlier, Wallach proposed that source elevation can be cued by the rate of change in interaural relationships relative to the rate of change in head orientation. It turns out that listeners are able to detect differences in the velocity of an auditory target with virtually the same degree of accuracy as they can detect differences in the velocity of a visual target (Waugh, Strybel & Perrott, 1979). It seems likely that visual detection of the rate of head movement would require the pick-up of similar information to that for the visual detection of the velocity of a moving visible target. Waugh et al. (1979) point out that their finding suggests that velocity judgement may involve a central process common to both modalities. If the elevation of a source can be judged on the basis of integration of information from both the visual system and the auditory system concerning rates of change, as Wallach's theory suggests, a central process would seem necessary.

### 1.2.5. Other empirical evidence for a role for head motion in sound localization

Somewhat surprisingly, other empirical research has provided only limited support for Wallach's (1939; 1940) theory. Young (1931) investigated the role of head movement by testing listeners' ability to localize sounds under conditions that eliminated any possible effect of head movement. To accomplish this he used a 'pseudophone', which was essentially a pair of ear trumpets, mounted one head-width apart on a rigid stand. Each trumpet was attached to the end of a rubber tube. The tubes passed through a wall and were terminated with ear-pieces that were inserted into a listener's ears (left trumpet to left ear, right to right). While the listener sat isolated from the ear trumpets, clicks from a telephone receiver were presented from various positions around the ear trumpets. While fairly proficient in making right-left judgements, listeners were unable to make accurate directional judgements with regards to up-down and front-back dimensions.

Young suggested that because up-down, front-back localization was not demonstrated under the conditions of his experiment, head movements must account for the unambiguous localization occurring in everyday situations, and that very small head movements, of as little as 1°, would have played a role in the relatively proficient localization reported by other researchers (Ewart, 1930; Matsumoto, 1897; Starch, 1905, cited in Young, 1931). Because the role of the pinnae was not well understood at the time, Young did not appreciate that their directional effects would have had an important bearing on the outcomes of these early studies. Also, the use of the pseudophone would not only have rendered head motion ineffective but would have also prevented the function of listener's pinnae. A logical possibility is that Young's experimental outcomes may have occurred entirely as a result of the pseudophone's disruption of pinna cues. Therefore it cannot be concluded that head motion plays a role in auditory localization on the basis of Young's (1931) evidence.

Earlier, Young (1928) conducted a study that had the potential to provide clear evidence for a role of head movement in unambiguous sound localization. Indeed, Young (1931) advanced findings from this earlier study as evidence of such. In the earlier study (Young, 1928), listeners were permitted free head motion and at times encouraged to employ it. However, Young did not systematically record when head movement occurred and how the head was moved, so the evidence from this experiment is somewhat ambiguous. Young (1928) was investigating localization ability when left and right ear inputs are reversed. He attached a pseudophone on the listener's head, configured so that the left ear was connected to an ear trumpet mounted near the right ear and vice versa for the right ear. Thus, the interaural relationships were reversed. Young observed that listeners wearing this apparatus sometimes judged sources to be 180° from their actual position (taking into account the tranposition of left and right, this outcome would be predicted by Wallach's theory, if head rotation was employed, or by classical theory if a motionless listener had made a front-back error). On other occasions, listeners judged sources to lie symmetrically opposite the actual source with respect to the MVP (an outcome that might be expected if head rotation was not employed and front-back judgement happened to be correct). If head motion had been documented more precisely the evidence from such an experiment should have provided a clearer picture about the role of head motion.

Klensch (1948, cited in Blauert, 1983) and Jonkees and van der Veer (1958) carried out experiments that gave some support to Wallach's theory. They employed two stethoscope tubes, with one end of each being inserted into a listener's ear and the other connected to the narrow end of a metal funnel. The funnels were positioned, approximately head-width apart, in front of the listener with their wide ends facing a sound source. The right funnel fed sound to the right ear and the left funnel fed sound to the left. When the head and funnels were absolutely motionless, the sound could not be localized, with the listener experiencing a sound image within their head. When the head was rotated, oscillating between right and left, and the funnels were moved forward and backward, coincident with the movement of the ears, the source was correctly localized as being in front. Transposing the left and right funnels so that the tubes crossed over, or moving the funnels in the opposite directions to the movements of the ears, gave listeners the impression that the sound was produced from behind. Koenig (1950) observed similar results with two microphones attached to an artificial head. The signals from the microphones were separately amplified and fed to the listener's ears. When the artificial head remained motionless, front-back discrimination was not possible, but when rotated in the same direction as the listener's rotating head, front-back localization was accurate. Although such experiments strongly suggest a role for head movements in front-back discrimination, the conditions in which the experiments were conducted are not entirely representative of conditions occurring in the everyday world.

Burger (1958) tested listeners on a front-back discrimination task with each of 12 different octave bands of white noise. Each band represented a different frequency region of the acoustic spectrum from 150 Hz to 12.8 kHz. The experiment included conditions in which: 1) head movements were permitted, although the listener was asked not to make excessive head movements; 2) head movements were permitted but listeners wore non-functional earphones over their ears to prevent pinna function; and 3) the head was tightly clamped to prevent any head movement. Results showed that without pinna cue disruption, front-back discrimination was slightly more accurate when head movements were permitted, as opposed to when the head was tightly clamped, but substantial front-back confusion was still evident. In the free head motion condition, involving pinna cue disruption, front-back discrimination was about as proficient as the clamped head condition for low-frequency octave bands, but substantially less proficient for the high-frequency bands. These results suggest that head movement is not as effective in disambiguating source position, under conditions likely to occur in the everyday world, as Wallach's theory implies. However, as with the study of Young (1928), it is not clear that Burger's listeners moved on every 'movement permitted' trial.

It should also be noted that Burger's results suggested that when the pinnae were not covered, they were able to function, in spite of the reportedly limited range of frequencies present within each of the bands. Current understanding of pinnae function would suggest that the filtering employed by Burger could not have restricted the signal to just one octave, at least for those bands occupying frequency ranges below 4 kHz. Yet when head movements were permitted, the condition with unnocluded pinnae produced more accurate front/back discrimination than the condition with occluded pinnae, no matter which octave band was involved.

Pollack and Rose (1967) tested localization of broad-band signals in the front HP and concluded that head movement contributed little except when signal duration was at least 3 s, and sources were at extreme lateral positions. Their conclusion was not well founded since they restricted testing to visible sources in the front HP, therefore front-back—up-down localization was untested.

Thurlow and Runge (1967) tested localization of 5-second bursts of low- and high-frequency noise presented from loudspeakers occupying seven general directions, in an experiment that involved various types of mechanically induced head movements. Their results suggested that head rotation of 45° eliminates front-back confusion, although that conclusion may be questioned. Listeners were tested with sources occupying a limited range of directions. Data from only one direction, approximately directly to the front at 41° elevation above the HP, was used to assess front-back discrimination. The results showed that when listeners did not move during sound presentation, the frontal source was misattributed to behind, by 90% of listeners, when the noise was low-frequency, and by 20% when the noise was high-frequency. Yet if their head was rotated through 45°, no listener attributed the low- or the high-frequency noises to behind; an outcome consistent with the hypothesis that front-back confusion is eliminated with head rotation. However, because the experiment did not involve a sound source positioned behind the listener, at a similar azimuth to the frontal loudspeaker, it is possible the results of the experiment reflect a response bias to the front activated whenever 45° of mechanically induced head rotation was involved. Thus, listeners may not actually have been able to discriminate between front and back.

From the same experiment Thurlow and Runge (1967) also obtained results suggesting that with low-frequency noise, head movements produced a significant but very slight increase in accuracy of vertical localization. Again the results may have been affected by experimental limitations. Thurlow and Runge did not test localization of sources elevated more than 41° from the HP and conducted testing with listeners under blindfolded conditions.

Fisher and Freedman (1968) tested the front-back discrimination ability of blindfolded listeners using trains of 40-ms pulses of white noise presented from loudspeakers spaced 45° apart throughout the HP. Listeners were tested in conditions with their pinnae able to function normally; with tubes inserted into their ear canals, disrupting pinnae function; and with artificial pinnae, made from casts of real pinnae, attached to the ends of the tubes inserted into their ear canals. In all these conditions, the same general finding showed that front-back discrimination was highly accurate in conditions allowing free head movement, but substantially inaccurate when listeners remained motionless. Other studies (e.g., Thurlow & Runge, 1967; Thurlow & Mergener, 1970) have also provided evidence that clearly suggests that localization is more accurate when free head movement is employed. Thurlow and Mergener (1970), using the same loudspeaker positions as Thurlow and Runge (1967), concluded that free head movement assists localization of low-frequency noise when the signal is at least 1 second in duration, while performance approaches an optimal (although not very accurate) level with 2-second signals. However, it is not clear whether the observed superior front-back localization depended on head motion cues, as outlined by Wallach, or merely on more efficient use of static localization cues once the head is reoriented.

Limitations of experiments designed to assess the contribution of head motion mean that the role of such movement for sound localization remains unclear. As noted by Middlebrooks and Green (1991, p. 153):

> *In light of all the evidence, a defensible argument is that unless the sound duration is sufficient to allow the listener to turn to face the source, thereby obtaining the optimum static localization cues,*

> *moving one's head may indeed be a poor strategy for improving the accuracy of localizing short-duration sources.*

In the following chapters, the results of nine experiments are reported, which were designed to investigate the effect of head motion on localization. By using actual sound sources, sighted listeners in normally lit conditions, and sounds providing localization cues that are ambiguous under static conditions, such experiments directly test various aspects of Wallach's theory in a way that equates with this auditory function in real world conditions.

# Chapter two: Preliminary experiments

This chapter reports on three preliminary experiments designed to investigate the role of head motion and to offer an initial test of Wallach's theory. The first experiment was somewhat modelled on that of Thurlow and Runge (1967), but with loudspeaker coverage from 0° to 90° elevation and without mechanically induced head motion.

## *2.1. EXPERIMENT 1*

### 2.1.1. Method

The first experiment involved 3 listeners (2 males, 1 female) all reporting normal hearing. They were presented with noise bursts from 19 loudspeakers (Realistic

Midrange Tweeters), selected from a larger set as being most closely matched in terms of frequency response. These, positioned at 15° intervals, formed two arcs, one spanning the left lateral HP from directly in front to directly behind, the other spanning the upper-left quadrant of the LVP from directly left (on the interaural axis) to directly above. All loudspeakers faced towards the centre of the listener's head at a distance of 1.25 m and were mounted on a frame constructed of 2-cm box-section, steel tubing. The apparatus was housed in a sound-isolated, semi-anechoic test room, 3 m wide × 3.5 m long × 3 m high.

To achieve a low level of constraint on listeners' responses, the actual whereabouts of the loudspeakers was visually masked, as shown in Figure 5, by the use of a partial spherical screen of 1-m radius. This comprised a fine-weave acoustically transparent fabric suspended inside a framework of hoops constructed of 1-cm diameter steel rods. Response options were marked on the inside surface of the fabric and took the form of irregular two-term, letter-number strings. These occurred at positions 15° apart in terms of elevation and azimuth — as defined by the double-pole co-ordinate system used by Middlebrooks, Makous and Green (1989) and reproduced here as Figure 4 (Chapter 1, page 13). Except for openings in the screen allowing for an entrance and sitting position, response options were available throughout spherical space. The screen was positioned such that there was a response option marked on its surface in

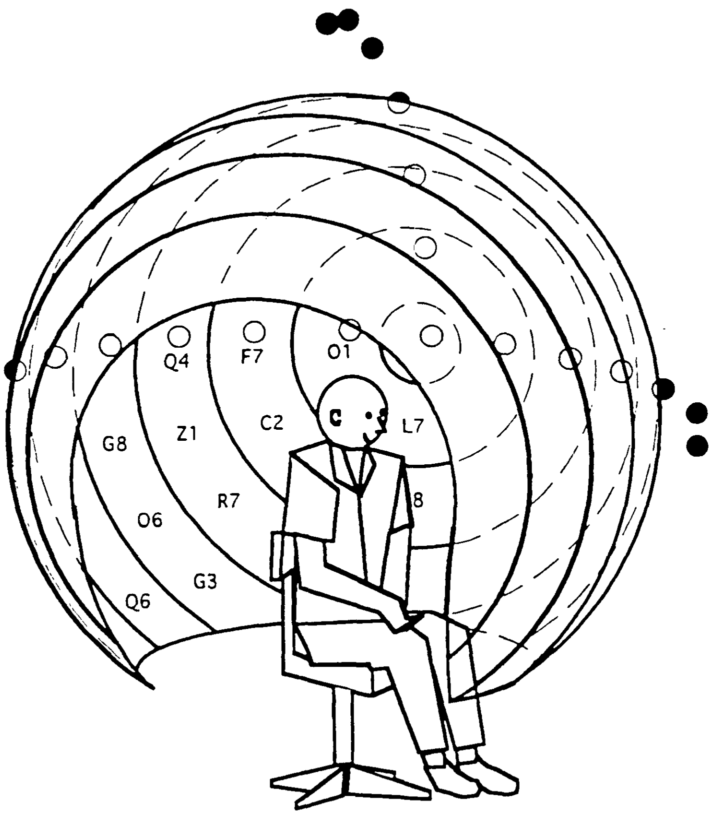

Figure 5. Loudspeaker layout relative to partial sphere structure. A sample area of the co-ordinate marking of the interior surface is shown. Co-ordinate labels use irregular two-term letter-number strings; co-ordinate positions occur at 15° intervals of azimuth and elevation.

alignment with each of the loudspeakers, as projected from the centre of the listener's head. Lighting conditions, created using two 40 watt incandescent lights, prevented listeners from seeing the actual loudspeaker positions through the screen. Additional fabric was suspended between the openings in the partial spherical screen and the floor and walls of the test room to complete the visual concealment of the loudspeakers.

Noise bursts comprised equal energy per 1/3 octave from 250 Hz to 2 kHz. This signal was obtained by passing randomly generated noise through a graphic equalizer (adjusted to remove the mean loudspeaker transfer function), followed by a 2-kHz digital low-pass filter with a rejection slope of ≥72 dB/octave. On each trial the signal was presented continuously for 3 s including 10-ms rise and fall times. While the mean level of the noise bursts was 60 dBA, the actual level was randomly adjusted from trial to trial by multiples of .375 dB over a range of ±5.25 dB, to prevent listeners from identifying loudspeaker positions based on minor level differences.

Listeners were required to localize the noise burst sources in each of 5 different conditions: 1) rotating the head from side to side about a central vertical axis; 2) pivoting the head from side to side about a central front-to-back horizontal axis; 3) tipping the head up and down about the interaural axis; 4) remaining perfectly motionless and 5) moving the head naturally, as they normally would when

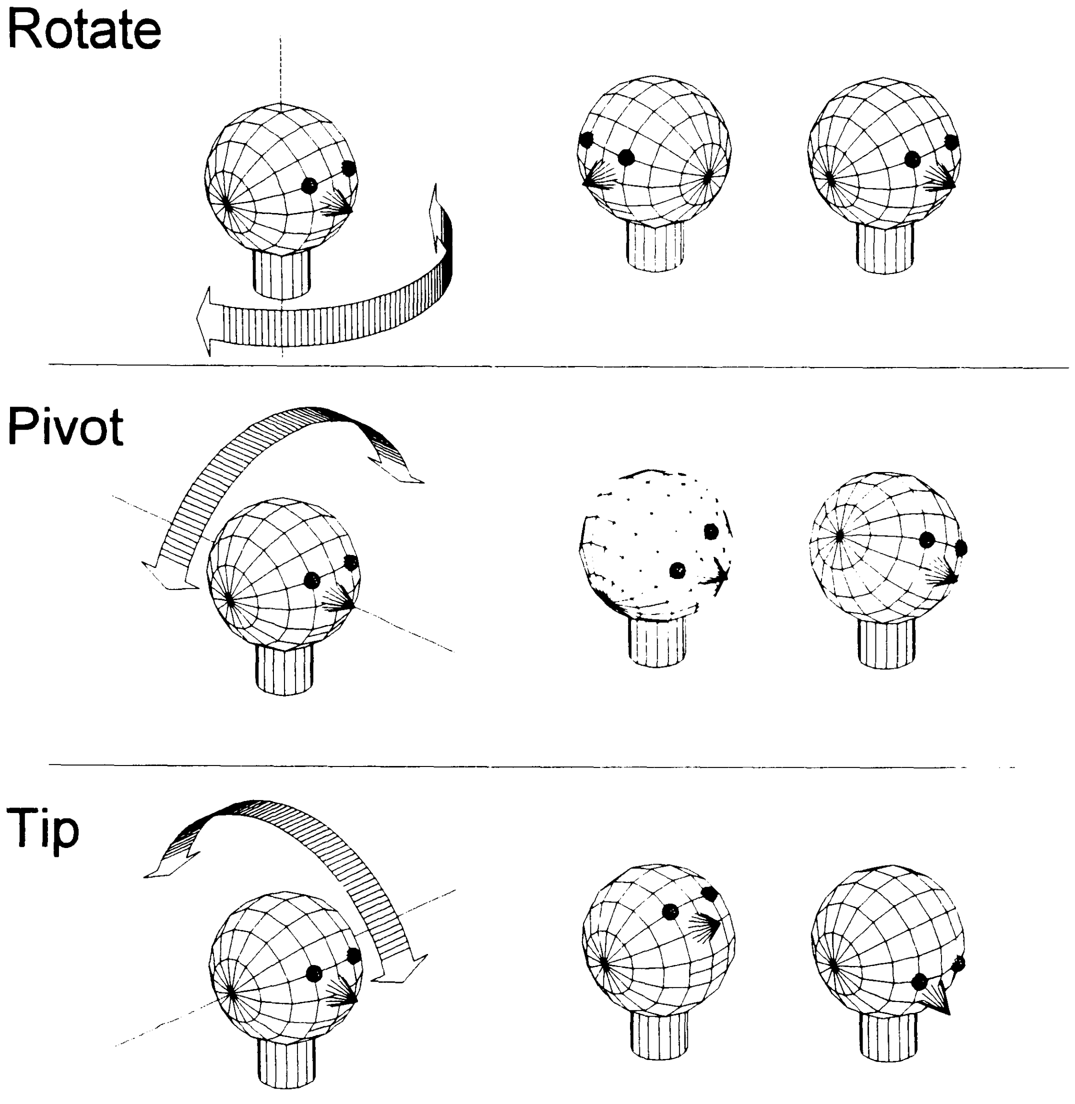

Figure 6. Three types of head movement: rotate, pivot and tip. For each movement type, a curved arrow indicates direction of movement and a straight line passing through the head indicates the axis of movement.

searching for the source of a sound (see Figure 6 for the first three types of movement). Conditions were completed in the order described. At the start of each trial, listeners were required to sit still, facing straight ahead, and wait for a noise burst to occur. In response to the onset of the noise burst, listeners were obliged to move or remain motionless, as the condition required, until signal offset. They could then move as they wished in order to report verbally the response option that best approximated the apparent direction of the source. Closed circuit TV monitoring was used throughout to check compliance with each condition. For each listening condition there were seven practice trials (to familiarize listeners with the specific actions required) followed by 19 experimental trials (one per loudspeaker position). No feedback on accuracy was given.

### 2.1.2. Results

During the running of the experiment it was revealed that the listeners found the pivot movement difficult to perform. Inspection of video tapes showed that listeners could not pivot their heads without also introducing small but clearly noticeable rotation movements. The results for the pivot condition should therefore be interpreted in light of this.

**Front-back errors.** A front-back error was recorded when a noise burst presented from a source in one hemisphere, forward or rearward, was erroneously attributed to a location in the opposite hemisphere. Trials involving sources positioned in the LVP were excluded from front-back error analysis. Table 1 shows that no front-back errors occurred in the rotation or natural movement conditions. Front-back errors occurred at the rate of 5.6% in the pivot condition, 19.4% in the tip condition and 22.2% in the motionless condition.

Table 1. Front-back errors for each condition in Experiment 1

| Rotate | Pivot | Tip | Motionless | Natural |
|---|---|---|---|---|
| 0/36 | 2/36 | 7/36 | 8/36 | 0/36 |
| (0%) | (5.6%) | (19.4%) | (22.2%) | (0%) |

Note: All were front-to-back errors

**Apparent elevation.** Apparent elevation is simply the elevation of a listener's response and is relied on for graphically representing listeners' elevation judgements. In Figure 7, it can be seen that the natural condition represents the most accurate localization of source elevation. In that condition, LVP and HP positions were localized at elevations close to their actual source elevations. In the other conditions, LVP and HP sources were generally localized within 30° of

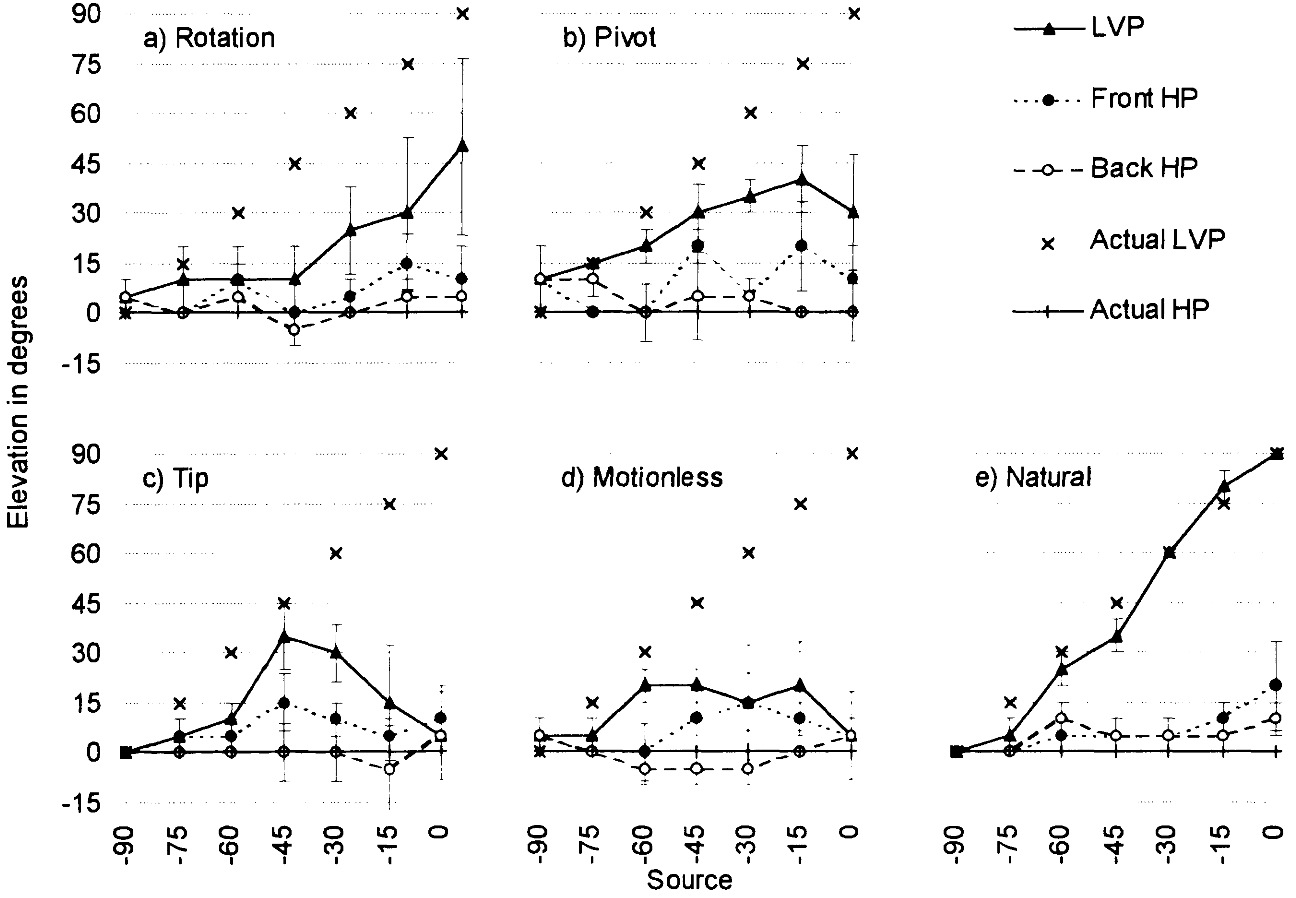

Figure 7. Apparent elevation of sources in the three quadrants tested in Experiment 1. Individual sources within the quadrants are identified in terms of their (double-pole) azimuth positions in degrees, from -90 (-90°, directly left) through to 0 (0°, in the MVP). Actual elevations are indicated by crosses (×). Error bars indicate standard error of the mean.

0° elevation. There were exceptions to this; most notably, for the source directly overhead, which was localized substantially more accurately in the rotation condition than in the motionless condition. Inspection of individual data for rotation revealed that two of the three listeners localized the overhead source within 30° of its true elevation. Compared to motionless listening, the pivot movement provided elevation judgement that was slightly more accurate for LVP sources in or near the MVP. The tip condition produced accurate responding for the LVP source at -45° azimuth but not for other LVP sources.

### 2.1.3. Discussion

These results show that natural head movement allowed more accurate localization of 3-s low-pass noise bursts than motionless listening. With natural movement, front-back errors were eliminated and elevation judgement was highly accurate. Head rotation eliminated front-back errors and, for some listeners, served to increase elevation accuracy for the source directly overhead. This outcome provides support, if limited, for Wallach's theory about the cue value of the rate of change in lateral angle relative to the degree of head rotation.

Results for the pivot movement show substantially fewer front-back errors and slightly greater accuracy in elevation judgement than for motionless listening. According to Wallach's theory a purely pivoting movement would not enable disambiguation of front and back. It may be that the small rotation movements listeners unavoidably made as they attempted to pivot were responsible for the substantial reduction in front-back errors. If this is so it would be expected that mechanically controlled pivoting of the listener's head, as employed by Thurlow and Runge (1967), would not produce a reduction in front-back errors. However, those experimenters obtained results for a low-band noise presented from in front which showed that 90% of listeners made front-back errors in the no motion condition, and that this reduced to 60% in the pivot condition. If it were certain that a mechanically controlled pivot movement was free of useful amounts of other movement, the conclusion would have to be made that a pure form of the pivot movement somehow provides listeners with information not formulated in Wallach's theory. On the other hand, it may be that the apparatus used by Thurlow and Runge to control pivoting did not entirely eliminate other forms of movement and that occasionally there was enough head rotation to remove front-back ambiguity. If so, the results of that and the present experiment suggest there need only be minimal rotation for front-back resolution.

In the present study, the pivot movement gave only a slight increase in accuracy of elevation judgement, compared with motionless listening. One reason for such a small increase in accuracy may be that the transformations of interaural differences resulting from a relatively pure form of the pivot movement do not normally occur. The pivot movement was difficult to perform, suggesting it is not a very natural type of head movement, at least for the people taking part in this experiment. For most positions, the tip condition largely failed to change localization accuracy, which is in line with Wallach's theory since such a movement does not produce a perceptible change in the interaural differences arising for any of the source positions use in this experiment. The accurate judgement of elevation for the LVP source at -45° azimuth suggests listeners were able to gain something from the tip movement. Again, the idea that movement can provide information not formulated in Wallach's theory must be considered. It is possible that the minor alterations in low-frequency spectra, produced through the signal's interaction with the shoulders and torso (Kuhn, 1987), provide some information about source elevation, and that this information is obtained more effectively if head motion occurs.

Signal duration in this experiment exceeded or matched that which earlier studies have shown to allow any contribution of head movement. Middlebrooks and Green (1991) make the point that there is basically no evidence for head motion having any effect on localization unless the signal is of sufficient duration to allow listeners to face towards the source. Once thus oriented, they are able to make use of optimal static cues. This would suggest that head motion may have no effect on localization of short duration signals. To address this point, the second experiment was designed to gauge the contribution of head motion to localization of sources of much shorter duration sounds.

## *2.2. EXPERIMENT 2*

### 2.2.1. Method

The second experiment involved four listeners (all females) reporting normal hearing and with no previous experience in localization experiments. The technical details were the same as before except that the signal duration was 0.5 s. The procedure was also the same as before, with five movement conditions: natural, rotation, pivot, tip and motionless, executed in that order. It appeared that natural movement did not occur before signal offset, so the natural movement condition was replaced with a block of motionless trials for two of the listeners.

## 2.2.2. Results

Only two listeners participated in a natural movement condition, and observations of their natural head motion revealed that movement did not occur until the 0.5-s signal had actually or virtually ceased. Thus results for that condition in this experiment are not considered (a 0.5-s natural condition was included in a subsequent, larger-scale experiment — Experiment 4 — for which the results will be reported).

**Front-back errors.** Table 2 shows that the only condition in which front-back errors did not occur was that of rotation. Only 2.1% of HP trials resulted in front-back errors in the pivot condition. The tip and motionless conditions produced front-back error rates of 22.9% and 20.8% respectively.

Table 2. Front-back errors for each condition in Experiment 2

| Error type | Rotate | Pivot | Tip | Motionless |
|---|---|---|---|---|
| Front-to-back | 0/24 | 0/24 | 9/24 | 7/24 |
| Back-to-front | 0/24 | 1/24 | 2/24 | 3/24 |
| All | 0/48<br>(0%) | 1/48<br>(2.1%) | 11/48<br>(22.9%) | 10/48<br>(20.8%) |

**Apparent elevation.** The apparent elevation data presented in Figure 8 show very little difference between any of the conditions. Greatest accuracy occurred in the pivot condition with the apparent elevation of LVP sources clearly greater than that for HP sources. Accuracy was slightly reduced in the rotation condition. Both pivot and rotation produced somewhat greater elevation accuracy than the motionless condition for sources in the front HP. The tip condition produced the least accurate responding, with the apparent elevation of LVP, FHP and BHP sources often falling between 15° and 30° above the horizon.

### 2.2.3. Discussion

The main finding from this experiment is that a long period of time is not required before head motion makes an important contribution to localization accuracy. As with the 3-s signal in the first experiment, rotation of the head completely eliminated front-back errors. Bearing in mind that listeners were instructed to remain motionless until they heard the onset of the sound before commencing their rotation movement, the period of time in which the head was rotating while the sound was occurring would have been considerably less than the 0.5-s signal duration. Once again the pivot movement also produced very few front-back errors and this may be due to the incidental rotation movements

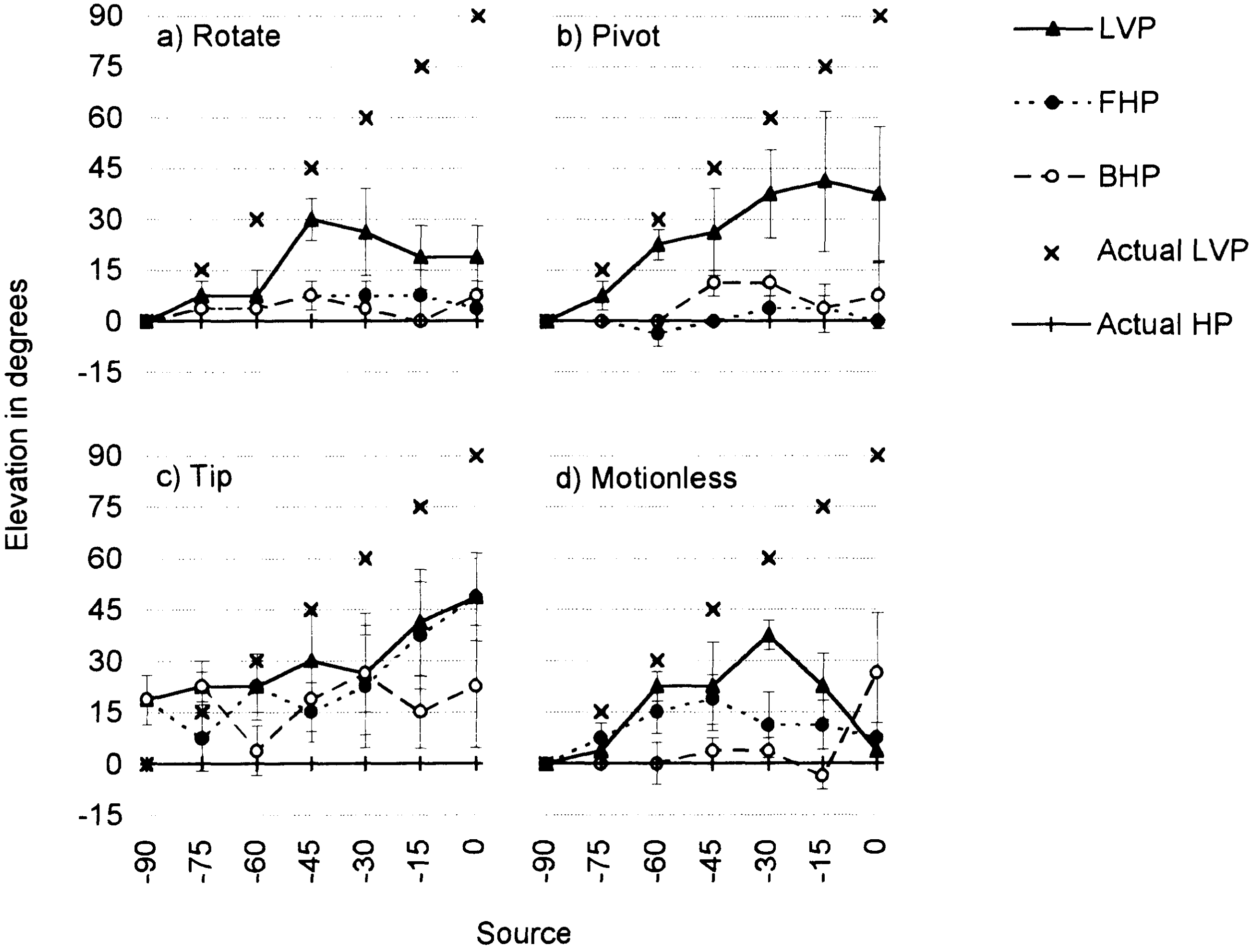

Figure 8. Apparent elevation of sources in the three quadrants tested in Experiment 2. Individual sources within the quadrants are identified in terms of their (double-pole) azimuth positions in degrees, from -90 (-90°, directly left) through to 0 (0°, in the MVP). Actual elevations are indicated by crosses (×). Error bars indicate standard error of the mean

occurring while listeners attempted the pivot movement, confirming the prospect that minimal rotation achieves front-back resolution.

In contrast to the results in the first experiment, elevation accuracy increased very little with rotation, with the apparent elevation of the source directly overhead being localized near the horizon. Although far from being proficient, the pivot movement produced the most accurate localization in terms of apparent elevation. The response pattern for the pivot condition was similar in both experiments. As in the first experiment, the tip condition produced highly inaccurate localization, in line with Wallach's theory.

The results from the first two experiments clearly suggest that listeners are able to localize sounds more accurately with head motion than without. These results were obtained with very few listeners, 3 in experiment 1 and 4 in experiment 2. Considerable variability between listeners was observed in both experiments even though all reported normal hearing. A signal lacking in the frequencies necessary for pinna function would make localization quite difficult. Without a clear sense of source direction, listeners' responses are likely to be erratic. Thus, to reduce experimental error, it was deemed necessary to conduct a further preliminary experiment with a larger sample size.

The contribution of rotation to localization of long and short duration signals was chosen as the focus of the third experiment because rotation had reliably eliminated front-back errors with both signal lengths and was readily achievable by listeners. Natural movement with a long duration signal was included since it represented the optimal level of performance achievable. Motionless listening was included as a baseline condition to compare with the movement conditions.

## *2.3. EXPERIMENT 3*

### 2.3.1. Method

The third experiment involved nine listeners (7 females, 2 males) all reporting normal hearing and with no previous experience in localization experiments. The loudspeaker set-up and signal were the same as before and involved the two different durations: 0.5 s and 3 s. Four conditions were executed in five separate blocks of 19 trials in the order listed: 1) rotation with a 0.5-s signal, 2) rotation with a 3-s signal, 3) Natural movement with a 3-s signal, 5) rotation with a 0.5-s signal and 5) Motionless listening with a 0.5-s signal. In all the rotation blocks, listeners were asked to make a single rotation of the head to face left in response to the onset of a noise burst and to make no other movements until the offset.

This single rotation movement was used so that greater uniformity between listeners could be assured. In the natural movement condition listeners were free to move in the way they normally searched for the source of a sound, as soon as they heard the onset of the sound. In the motionless condition, they were required to remain motionless throughout the signal. All other procedures were the same as in the first two experiments.

## 2.3.2. Results

**Source-head-response angle.** A new measure of absolute accuracy was devised, and is expressed as the source-head-response (SHR) angle — the angle describing the relation between the position of the source, the centre of the listener's head, and the position of the response, on any trial. Table 3 displays the resulting mean SHR angle for each condition, showing that the mean error was significantly larger in the motionless condition than in the rotation or natural conditions. None of the other differences was significant.

**Front-back errors.** Only the motionless condition produced front-back (front-to-back and back-to-front) errors. These occurred on 17.6% of motionless

Table 3. Source-Head-Response angle means and standard deviations for each condition in Experiment 3

| Order of testing | Signal duration | Movement | Mean | Std. Dev. |
|---|---|---|---|---|
| 1 | 0.5-s | Rotation | 22° | 6.2° |
| 2 | 3-s | Rotation | 17° | 6.4° |
| 3 | 3-s | Natural | 16° | 6.4° |
| 4 | 0.5-s | Rotation | 21° | 7.5° |
| 5 | 0.5-s | Motionless | 35°* | 10.8° |

*Significantly different (p < .01).

Table 4. Front-back errors for each condition in Experiment 3

| Error Type | 0.5-s Rotation (time 1) | 3-s Rotation | 3-s Natural | 0.5-s Rotation (time 2) | 0.5-s Motionless |
|---|---|---|---|---|---|
| Front-to-back | 0/53 | 0/54 | 0/54 | 0/54 | 12/54 |
| Back-to-front | 0/54 | 0/54 | 0/54 | 0/53 | 7/54 |
| All | 0/107<br>(0%) | 0/108<br>(0%) | 0/108<br>(0%) | 0/107<br>(0%) | 19/108<br>(17.6%) |

Note: Two front-back errors for 500-ms rotation conditions not included because listener failed to rotate head before signal offset.

trials, as shown in Table 4. These results are consistent for corresponding conditions in the first two experiments.

**Apparent Elevation.** The apparent elevation data presented in Figure 9 show that, in the 3 second rotation and natural conditions, listeners judged the elevation of sources with about the same level of proficiency as in the first experiment. It can be seen that 0.5-s rotation and motionless conditions differ very little; the only exception is that the elevation of the loudspeaker directly overhead was judged somewhat more accurately in the 0.5-s rotation condition than in the motionless condition. Table 5. shows individual performances for elevation judgements for the loudspeaker directly overhead. Some listeners appeared to gain good cues for the elevation of this source from head rotation. Even when the sound lasted for 0.5-s, four listeners at time 1, and three at time 2, made judgements for the overhead source that were within 30° of the correct elevation, whereas no listener could do this in the motionless condition.

## 2.3.3. Discussion

This experiment clarifies the earlier outcomes, and shows Wallach's theory may be supported to some extent. Head rotation clearly assisted listeners so that they

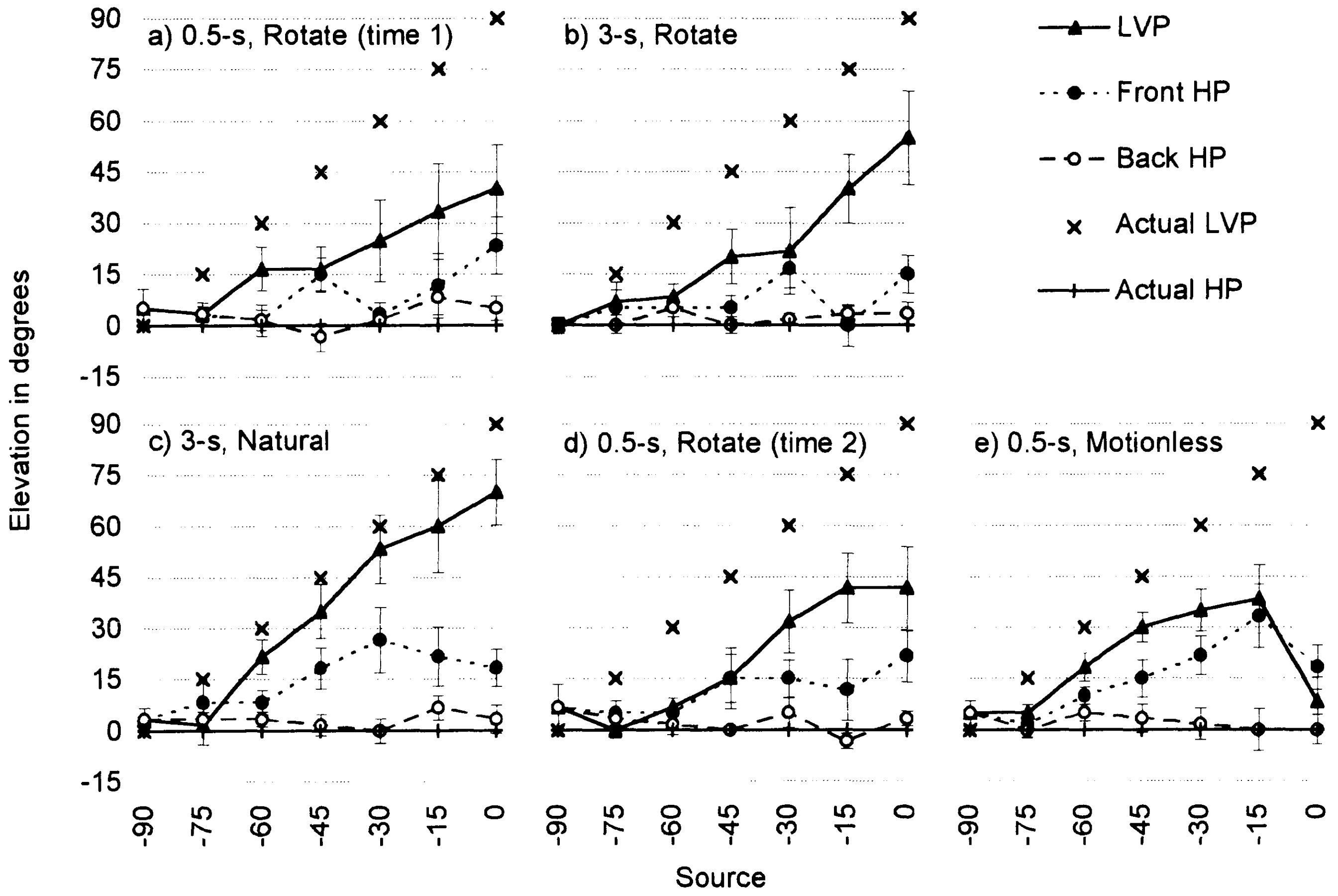

Figure 9. Apparent elevation of sources in the three quadrants tested in Experiment 3. Individual sources within the quadrants are identified in terms of their (double-pole) azimuth positions in degrees, from -90 (-90°, directly left) through to 0 (0°, in the MVP). Actual elevations are indicated by crosses (×). Error bars indicate standard error of the mean.

Table 5. Elevation error occurring in Experiment 3 for loudspeaker situated directly over listeners head. Boxes indicate responses where margin of error was ≤ 30°.

| Listener No. | 0.5-s Rotation | 3-s Rotation | 3-s Natural | 0.5-s Rotation | 0.5-s Motionless |
|---|---|---|---|---|---|
| 1 | $\boxed{0^\circ}$ | -135° | $\boxed{0^\circ}$ | -90° | -120° |
| 2 | -90° | $\boxed{-30^\circ}$ | -90° | -90° | -90° |
| 3 | -90° | -60° | $\boxed{0^\circ}$ | $\boxed{-15^\circ}$ | -90° |
| 4 | -75° | $\boxed{0^\circ}$ | $\boxed{-15^\circ}$ | -75° | -75° |
| 5 | $\boxed{-15^\circ}$ | $\boxed{-15^\circ}$ | $\boxed{0^\circ}$ | $\boxed{0^\circ}$ | -90° |
| 6 | $\boxed{0^\circ}$ | $\boxed{-30^\circ}$ | $\boxed{0^\circ}$ | $\boxed{0^\circ}$ | -90° |
| 7 | -60° | $\boxed{-15^\circ}$ | $\boxed{-15^\circ}$ | -45° | -60° |
| 8 | $\boxed{-30^\circ}$ | $\boxed{-15^\circ}$ | $\boxed{-30^\circ}$ | -75° | -45° |
| 9 | -90° | $\boxed{-15^\circ}$ | $\boxed{-30^\circ}$ | -45° | -75° |

did not make any front-back errors. The contribution, in terms of judging elevation, seems to be limited to localization of sources situated directly overhead.

The third experiment shows that head movements contribute significantly to localization of sources of low-pass noise even when the duration of the signal is very short. Analysis of video tapes for the 0.5-s rotation conditions revealed that the mean duration under which head rotation and noise burst occurred simultaneously was 260 ms for time 1 and 283 ms for time 2, with listeners often not completing the 90° rotation before signal offset. This suggests that the auditory system is very sensitive to spatial cues brought about by head rotation. Head movement occurring for only a small amount of time while a sound is ongoing is enough to make head motion cues useful to the listener.

One way of interpreting the lack of front-back errors, in the 3-s rotation condition at least, need not rely on the existence of dynamic cues. This is because, after 90° rotation of the head, listeners were oriented such that they merely had to make a left-right discrimination of sources, easily achievable with static interaural cues alone, since sources initially to the front ended up to the right and those initially to the rear ended up to the left. This relates to the point made by Middlebrooks and Green (1991) who suggested that any benefit of head movement may be because such movement allows a listener to achieve a head orientation that enables use of optimal static cues. Although the 90° leftward rotation was not spontaneously enacted by the listener, thus the final head orientation may not have been appropriate for optimal static localization cues for generally, the final head position did provide optimal cues for eliminating the kind of errors (front-back) that occurred without movement. To adequately test for a dynamic front-back cue, further experimentation needed to take the possible use of optimal static cues into account.

The preliminary experiments provided evidence that head movements assist localization of sound in circumstances where listeners' responses were largely unconstrained, there being a large range of possible response options available. Perrett and Noble (1995) showed limitations on response choices can seriously affect outcomes of localization experiments. Although the set-up used in the preliminary experiments featured low levels of constraint, the openings in the partial spherical screen meant that there were non response options below -45° elevation and none further to the right than +45° azimuth. The results of the preliminary experiments therefore were not completely unconstrained.

According to Wallach's theory, head rotations about a central vertical axis would not permit listeners to distinguish sources positioned above the HP from those positioned below the HP. Had listeners been permitted to respond to positions directly beneath them, their patterns of responding may have been different. For example, Wallach's theory suggests that when an interaural difference of zero remains unchanging while the head is rotated about a central vertical axis, a sound source lies either directly above or directly below. Without any other type of movement the remaining ambiguity is not resolvable. If listeners in the preliminary experiments had been given the opportunity to respond to positions below them, they may have responded there instead of above when head rotation produced no or virtually no change to a zero interaural difference. It is important therefore that in order to gauge what localization cues are produced by head motion, listeners should be completely free in their response options. The experimental arrangement was thus radically reconstructed to allow for responding in all possible directions.

# Chapter three: The effect of head motion with minimal constraint on responding

This chapter reports on two experiments (experiments 4 and 5) designed to investigate the role of head motion in auditory localization using an extensively modified experimental set-up which enabled responding that was almost completely unconstrained. The partial spherical screen was replaced with one that completely enclosed the listener. In addition, head-tracking equipment was installed for the purposes of recording head movements and providing apparent source co-ordinates for any position on the screen. A laser pointer mounted on a light-weight head harness was incorporated to assist listeners in making their responses. Additional loudspeakers were installed so that localization of sources at various other locations could be tested. Finally, to control more precisely for artefacts, a system was developed for inverse filtering of signals, to compensate for the spectral characteristics of individual loudspeakers.

# *3.1. EXPERIMENT 4*

The fourth experiment investigated the effects of natural head motion, head rotation and motionless listening on localization of 3-s and 0.5-s low-pass noise signals when sources were positioned throughout the left HP and the left LVP. This enabled a truer test of discrimination across hemispheres when head movement is limited to rotation. The use of a 0.5-s natural condition was to re-visit the question of whether voluntary movement delivers information when the signal is of only brief duration.

## 3.1.1. Method

The experiment involved 12 people (5 females, 7 males), all of similar background, and reporting normal hearing. None had previously participated in sound localization experiments. They were presented with low-pass noise bursts from 25 loudspeakers (Realistic Midrange Tweeters) at 15° intervals, forming two intersecting arcs, one spanning the left horizontal plane (HP) from directly in front to directly behind, the other spanning the leftmost lateral vertical plane (LVP) from directly above to directly below (see Figure 10). The loudspeakers were mounted on curved frames, constructed of 2-cm box-section steel tube, at a distance of 1.25 m from, and facing the centre of, the listener's head. The apparatus was housed in the same semi-anechoic, sound-isolated room used in the previous experiments.

A 1.2-m radius spherical screen completely surrounded the listener. This was constructed from acoustically transparent fine-weave fabric suspended in a framework of hoops made of 2-cm-diameter PVC tubing. The PVC frame for the screen was designed so as not to obstruct any of the loudspeakers, which were arrayed just beyond it. A hinged section of the screen could be swung open to allow access. Inside, a rotatable seat, height adjustable, was mounted on a platform (shown in Figure 10) made of 2-cm box-section steel tubing and weldmesh, which allowed the portion of the screen lying underneath to be seen: the legs of the platform pierced the screen to connect with the laboratory floor. Two 40 watt incandescent lights mounted near the inside surface of the spherical screen provided conditions such that the loudspeaker locations were not visible to the listener. Additional screening outside the sphere also prevented participants from gaining knowledge of actual loudspeaker placement. As far as they were aware, potential sound sources could lie in virtually any direction.

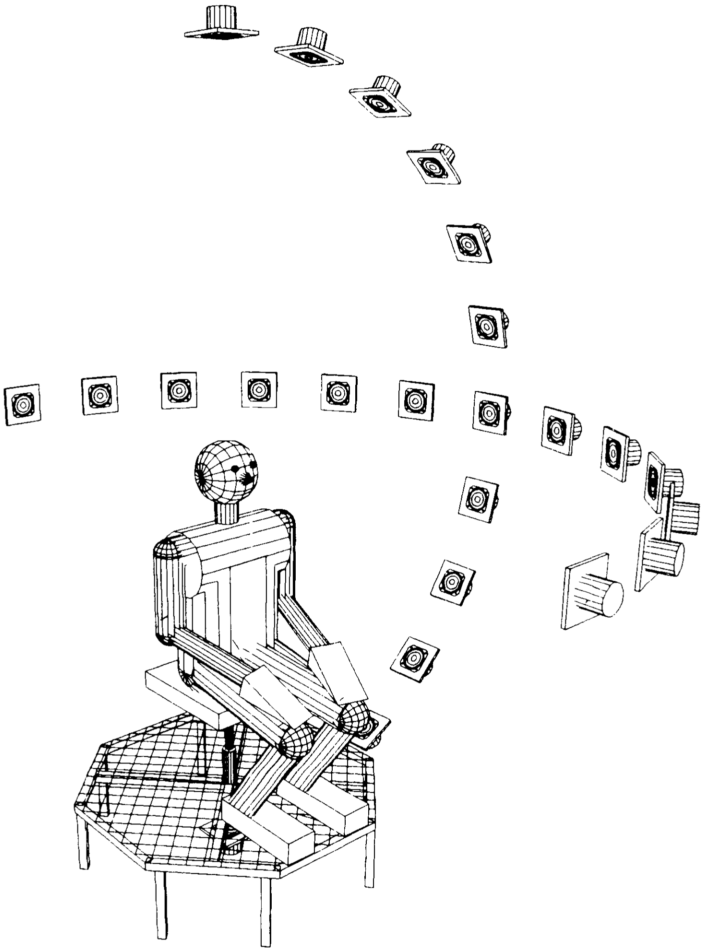

Figure 10. Loudspeaker array used in Experiment 4. Loudspeakers arranged in two arcs that intersect directly left of the listener who is seated on a rotatable seat supported by platform made of steel tubing and weldmesh. The spherical screen is not shown.

Polhemus Isotrak II head tracking equipment registered head position and orientation throughout each trial. The head tracker transmitter was positioned 480 mm to the right of the centre of the listener's head, mounted on the end of a length of 30-mm-diameter PVC tubing inserted through the screen. An unobtrusive adjustable head harness held the head tracker receiver and a laser pointer; the pointer was to assist listeners in making response decisions. TV monitoring was also used, the camera lens accommodated by making a small aperture in the screen.

The signal was created digitally, using the Matlab 'RANDN' function, to produce 3.08-second and 0.58-second samples of white noise (power density spectrum constant at all frequencies) with sampling rate of 44.1 kHz. Each sample was digitally filtered with a 255th order finite impulse response (FIR) filter to convert it to pink noise (equal energy per octave), then a 4096th order FIR low-pass filter to produce a 2.0-kHz cut-off with approximately 90 dB/octave rejection slope. Using custom written digital signal processing (DSP) software, each signal was inverse filtered (4096th order, FIR) to minimise effects of individual loudspeaker transfer functions. The start and finish of each sample was truncated by 0.04-s, to eliminate transients produced by the filtering process. To reduce onset/offset transients occurring at playback, the DSP software also applied a cosine squared windowing function with 20-ms onset/offset ramps. Thus, separate 3-second and 0.5-second samples were produced for each loudspeaker and stored as Microsoft

Windows WAVE format sound files on the hard disk drive of an 80386 DX/40 personal computer. The computer was fitted with a Sound Blaster 16 Vibra audio card which was used to convert the sound files into analog signals.

Custom written software running on a second (80286) computer, linked to the 80386, was used to control sound delivery during the experiment. The audio card analog output was passed through a custom built manually adjustable attenuator, used to set the mean overall signal level. Signals were then routed through a custom built computer-controlled attenuator which provided for random level variations, and also compensated for differences in individual loudspeaker efficiency. The signal was then amplified by a custom built power amplifier, and switched to one of the 25 loudspeakers via an array of self-cleaning relays; switching occurred approximately 100 ms before and after each sound delivery, thus preventing audible switching transients. The mean signal level was 55 dBA; the actual level was randomly adjusted from trial to trial in multiples of .375 dB over a range of ±3 dB to prevent identification of loudspeaker positions based on minor overall loudness differences.

Listeners were required to localize the sound sources in each of three movement conditions: 1) reacting as they normally would when searching for the source of a sound; 2) rotating the head 45° leftward on the HP, after the onset of the signal, and thereafter remaining motionless; 3) remaining motionless throughout. The choice of 45° rather than 90° for the rotation movement meant that listeners would be prevented from relying on static interaural cues for discriminating sources in the front HP from those in the back HP. Participation in each of the three movement/no-movement conditions was in two separate sessions, one with the 0.5-second signal, the other with the 3-second one. Movement and signal duration conditions were counterbalanced across participants.

Before each trial, listeners sat facing ahead, aiming the laser pointer at a dot of light, produced by small a light emitting diode (LED) positioned behind the screen, at 0° azimuth in the front HP. When ready, they pressed a button on a hand-held module which initiated head tracker data collection and, after a random delay of between 1 and 2-s, a noise burst was produced from one of the loudspeakers. In response to the onset of the noise burst, listeners moved or remained still, as the condition required. In the rotation condition, the head was rotated leftward until the pointer was aligned with a second, similarly produced, light-dot at -45° on the HP. At signal offset, in all conditions, listeners were permitted to move as they wished so as to aim the laser pointer where they judged the sound to have come from. Once satisfied that the pointer was aiming in the appropriate direction, the listener made a further button-push, which stopped head tracking data collection. The co-ordinates recorded by the head tracker at that moment, identified the judged direction of the source. For each listening condition there were seven practice trials followed by 50 experimental trials (two per loudspeaker position). The order of loudspeaker activation was random. No feedback on accuracy was given. From inspection of head tracker recordings, and from TV monitoring at the time of the experiment, it was evident that the different conditions were successfully complied with on almost every trial.

## 3.1.2. Results

**Source-head-response angle.** Average SHR angles in the six experimental conditions are shown in Table 6. A one-factor repeated measures analysis of variance revealed a significant difference across conditions, $\underline{F}$ (5,55) = 32.47, $\underline{p} < .001$. Generally, localization was more accurate when head motion occurred. Post-hoc testing (Tukey's HSD multiple comparisons) confirmed that the 3-second natural and rotation conditions produced significantly smaller SHR angles than the 3-second motionless condition ($p < .01$) and that the 0.5-second rotation condition produced significantly smaller SHR angles than either the 0.5-second motionless condition ($p < .01$) or the 0.5-second natural condition ($p < .05$). The SHR angles for 3-second natural and rotation conditions were not significantly different, nor were the SHR angles for 0.5-second natural and motionless conditions. The latter result arises because movement was less likely to be initiated in the 0.5-second natural movement condition prior to signal offset, hence that condition has features in common with motionless listening.

Table 6. Average SHR angle and absolute elevation error for each condition of Experiment 4 (standard deviations in brackets)

| signal (secs) | listening condition | SHR angle | absolute elevation error |
|---|---|---|---|
| 3 | Natural | 22 (6.6)** | 21 (5.5)** |
| 3 | Rotation | 26 (3.1)** | 23 (3.1)* |
| 3 | Motionless | 42 (7.1) | 28 (4.4) |
| 0.5 | Natural | 37 (8.3)* | 26 (4.6) |
| 0.5 | Rotation | 30 (4.1) | 25 (4.4) |
| 0.5 | Motionless | 41 (8.1)** | 28 (5.4) |

Tukey's HSD test, **p<.01; * p<.05

**Front-back errors.** Analysis of variance on percentages of front-back errors revealed significant differences across the six conditions, $F$ (5,55) = 20.61, $p < .001$. Figure 11 shows the percentage of front-back errors occurring in HP trials for each condition. Tukey's HSD tests established that the 3-second natural and rotation conditions produced significantly fewer front-back confusions than the 3-second motionless condition ($p < .01$) and that the 0.5-second rotation condition produced significantly fewer front-back confusions than either the 0.5-second motionless condition ($p < .01$) or the 0.5-second natural condition ($p < .05$). The percentages of front-back confusions for 3-second natural and rotation conditions were not significantly different, nor were the percentages of front-back confusions for 0.5-second natural and motionless conditions. Again, the latter result arises because movement is less likely to be initiated before the offset of the 0.5-second signal under the natural movement condition.

The head tracker co-ordinates at the moment of responding show there were a few front-back errors in the 3-second natural movement condition. These occurred with one listener, who opted not to move throughout the period of the signal. A few front-back errors occurred in natural and rotation conditions with 0.5-second signals, in cases where movement was less than 5° from the initial listening orientation. Other front-back errors occurred in the rotation and the 0.5-second natural movement conditions, even where movement was greater than 5°, but these were confined to sources in a spatial region around the interaural axis. Of the front-back errors occurring in the 3-s motionless conditions, 66% were front to back, and of those in the 0.5-s motionless condition, 55% were front to back.

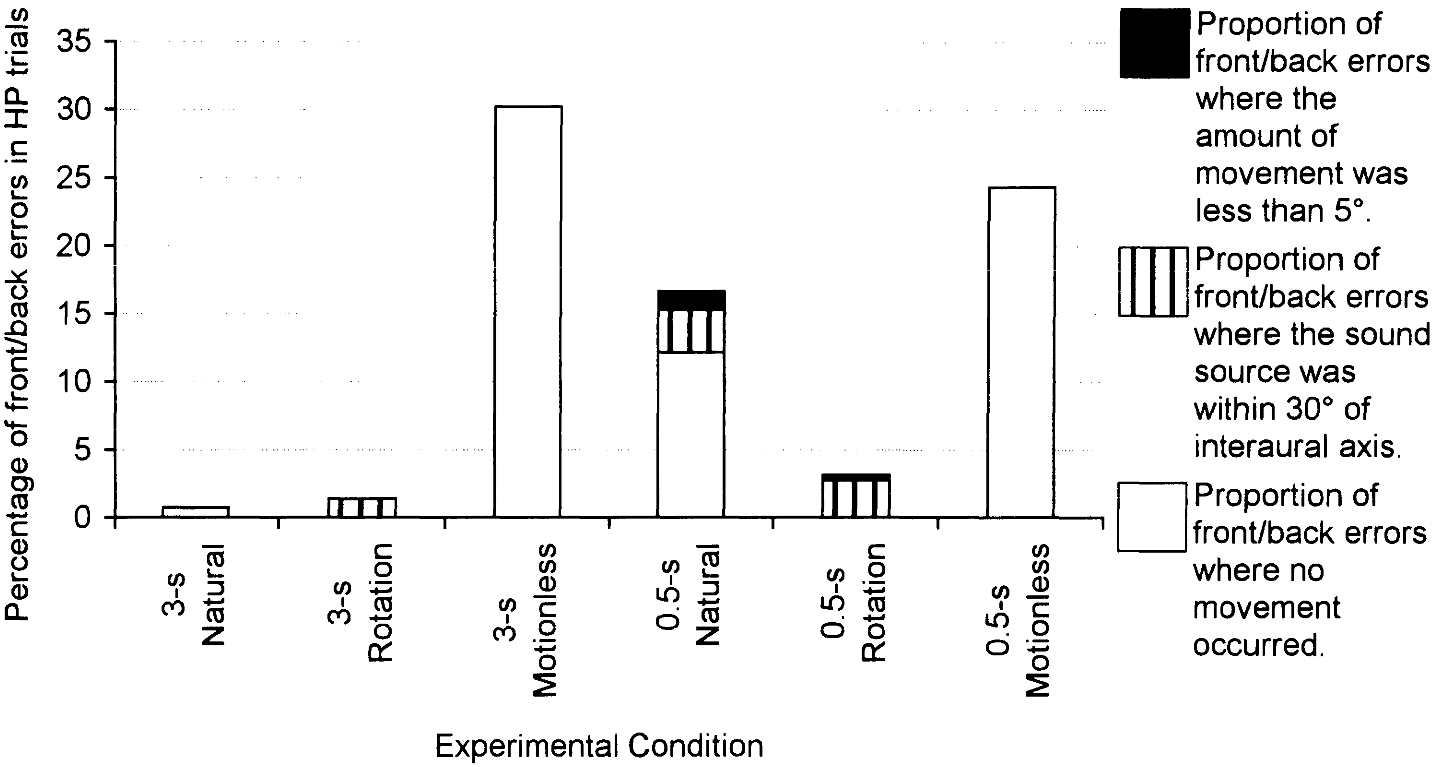

Figure 11. Front-back errors occurring in Experiment 4

**Absolute elevation error.** To assess the effect of movement on the elevation component of localization accuracy, analysis of variance was performed on absolute elevation error. Absolute elevation error is the vertical component of the SHR angle and is useful in statistical analysis of elevation judgements since it avoids the cancellation effect from adding error magnitudes of opposite sign (responses above and below actual positions) that occurs with signed elevation error. A significant difference in absolute elevation error was observed across conditions, $\underline{F}$ (5,55) = 9.71, $\underline{p} < .001$. Mean values are shown in the final column of Table 6 (on page 72). Post-hoc testing showed that the 3-second natural and rotation conditions produced significantly less error than the 3-second motionless condition ($\underline{p} < .01$ and $< .05$ respectively). There were no significant differences between the 3-second natural and rotation conditions nor among the various 0.5-second conditions.

**Apparent elevation.** The average apparent elevation of each sound source is shown in Figure 12 (3-second signal only) for sources in the LVP above the horizon, LVP below, HP in front of the interaural axis, and HP behind. It may be seen that natural movement offers little advantage over rotation in the perception of displacements above and below the horizon, although the upper LVP function is smoother for natural. Sources below the horizon, at least to a limit of -60° elevation (-30° azimuth), were distinguished in both movement conditions, and with a slight advantage over motionless listening in the -45° to -60° region of the

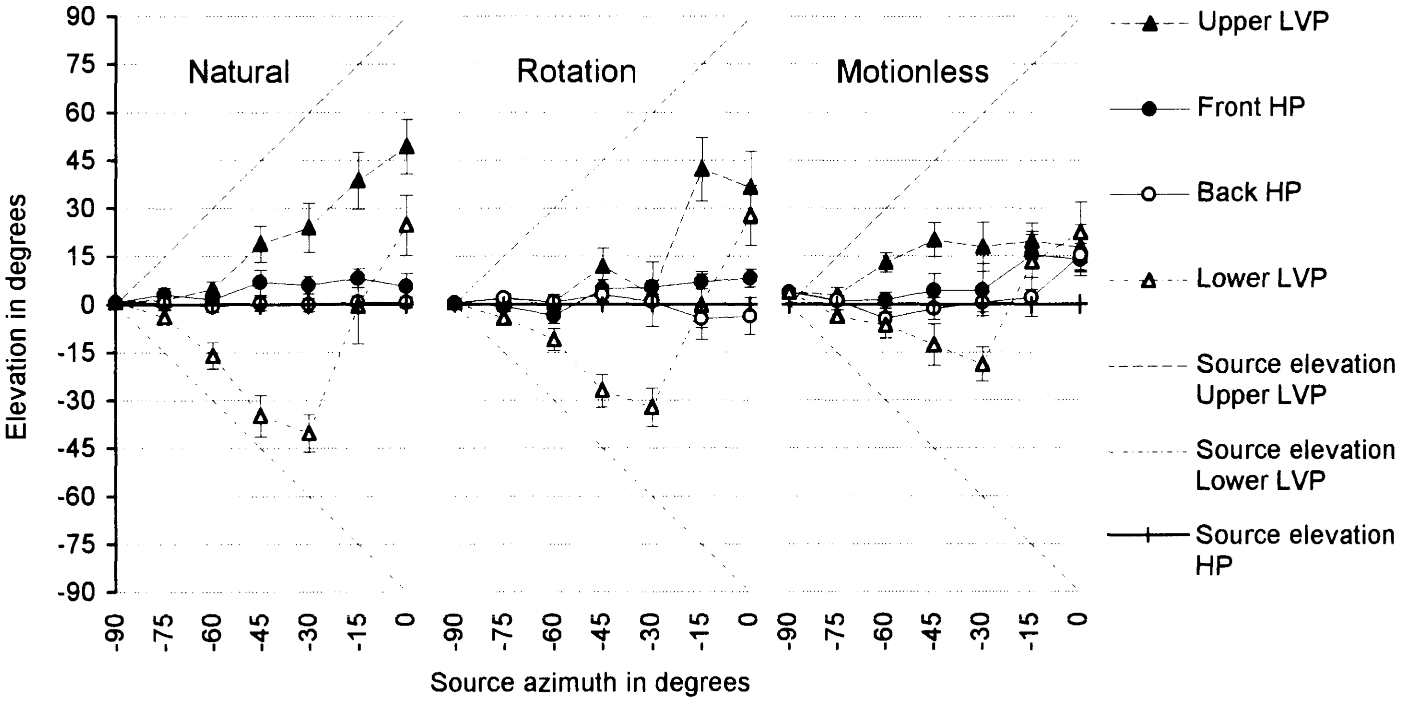

Figure 12. Apparent elevation of sources in the four quadrants tested in Experiment 4, under three test conditions for the 3-second signal. Error bars indicate standard error of the mean.

lower LVP. In all conditions there was a marked reversal of judged elevation for sources beneath the listener. From 0° to ±30° in the LVP, and throughout the HP, there were no differences among conditions. (For the 0.5-second signal there were also no differences among conditions across the range of locations in either plane, with patterns of responses in the upper LVP akin to those in the motionless condition shown in Figure 12, and patterns in the lower LVP akin to those in the rotation condition shown here.)

**Cluster analysis.** There were noticeable variations in the performance of different listeners, especially with respect to elevation judgement. Hierarchical cluster analysis was used to explore the existence of a typology of listeners based on absolute elevation error in each of the three movement/no-movement conditions with 3-second signals. Ward's minimum variance cluster analysis (Blashfield, 1976) was applied with dissimilarities between listeners' profiles being defined by squared Euclidean distance ($D^2$). The first large increment in aggregate $D^2$ occurred at the merging of three clusters into two, suggesting that three was an appropriate number to interpret. Clusters 1, 2 and 3 contained 5, 4 and 3 listeners respectively. Their performance patterns, in the form of average apparent elevation judgements, are given in Figure 13 a, b and c.

The first cluster showed an advantage for natural movement over rotation or motionless conditions in discriminating upper LVP positions; they showed no

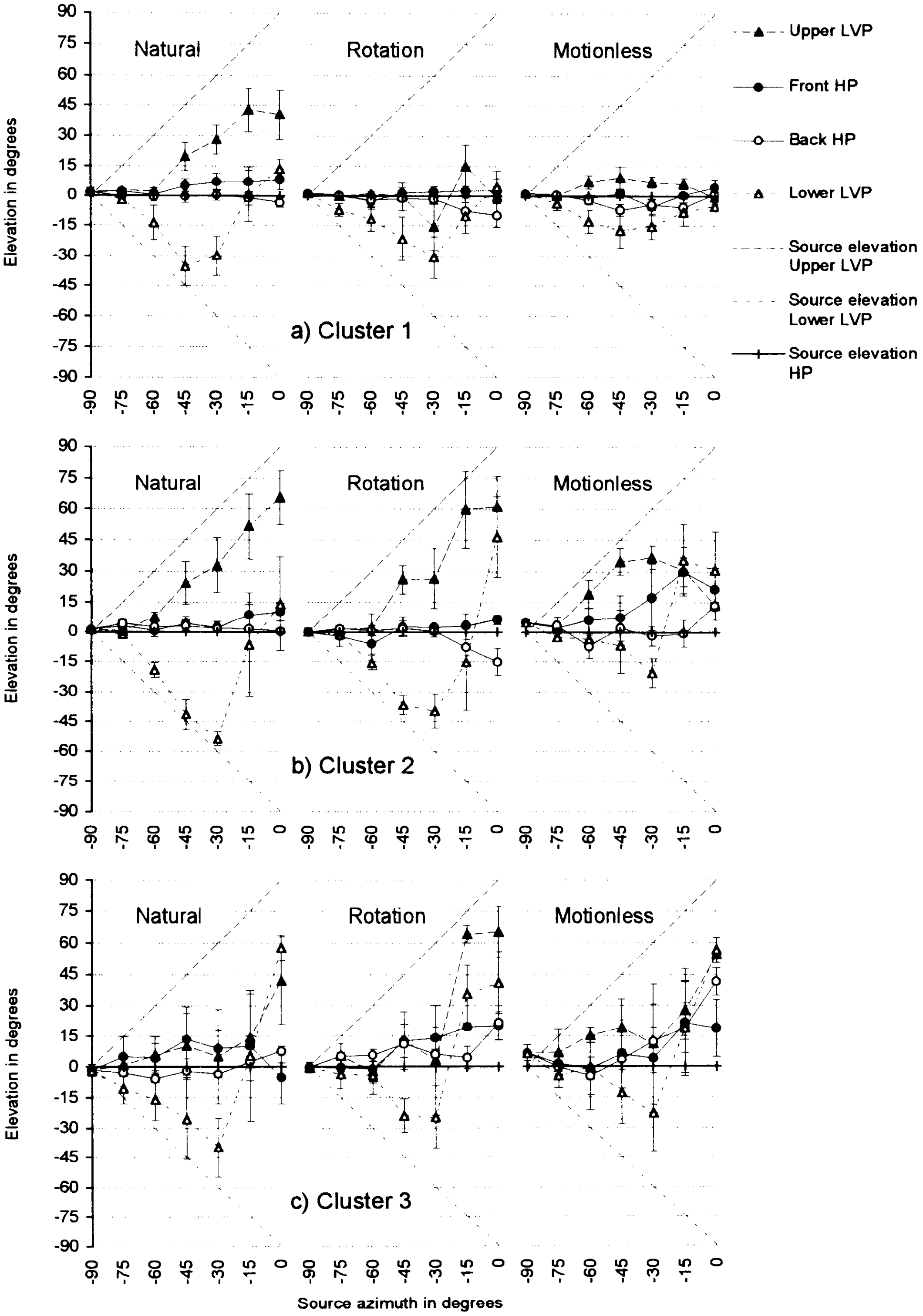

Figure 13. Apparent elevation in the same conditions as for Figure 12, but showing responses of the three cluster groups separately, in a, b and c respectively. Error bars indicate standard error of the mean.

clear performance difference between movement conditions in the lower LVP, and virtually no deviation from the horizon for sources on the HP. This relates to a point that cluster 1 showed little attribution of lower LVP sources to locations above the horizon.

The second cluster showed advantage for both natural and rotation movements in maintaining fairly proficient performance in the upper LVP, compared with motionless listening. There was a substantial attribution of the lowermost LVP source to above the horizon. An 'upward' bias in this group may partly explain the higher average apparent elevation for upper LVP sources in this cluster compared with cluster 1. The third cluster showed an advantage for rotation over both natural movement and motionless listening, in the uppermost region of the upper LVP. There are signs that the source overhead was heard as at or toward that location in all conditions for this cluster, which also showed a distinct attribution of the lowermost source to uppermost positions.

An examination of responses to the overhead source showed two distinct patterns, following instructed or natural movement: either there was a sense of this source being within ±45° of the horizon; or there was a sense of it being at least 60° above the horizon. Using +60° elevation as a criterion, it is noted that three of the 12 listeners located the overhead source correctly on each of the four 3-second movement trials they underwent with respect to that source; two others were correct both times in the rotation condition, and two others were correct both times under natural movement conditions. Of the five listeners achieving success in natural movement, one was successful both times under those conditions with the 0.5-second signal and three others were successful once under those conditions. Rotation with the 0.5-second signal led to success for one listener on one trial. On trials where the source was not detected as overhead, the head tracker records showed that a common response was to identify its whereabouts as either directly in front of or directly behind where the listener was facing at the point of signal offset.

Inspection of the forms of movement made in the natural condition (3-second signal) showed that, in response to HP signals displaced from the MVP, the typical first phase of movement was a single sweeping leftward rotation. In many cases this overshot the actual source location, especially in response to sources at -15° and -30°, and was followed by a return to the true position or a rapidly damped oscillation about it. In other cases, more noticeably in the region around the interaural axis and the back HP, a rapid initial leftward rotation was slowed before the true position was reached. When the source was directly in front there was either no or virtually no movement initiated, or there was an up-down tipping (nodding) of the head, or an oscillatory rotation. In the case of LVP sources at more than 45° from the HP, rotation was often combined with substantial downward and/or upward tipping of the head. Often the tipping movement began

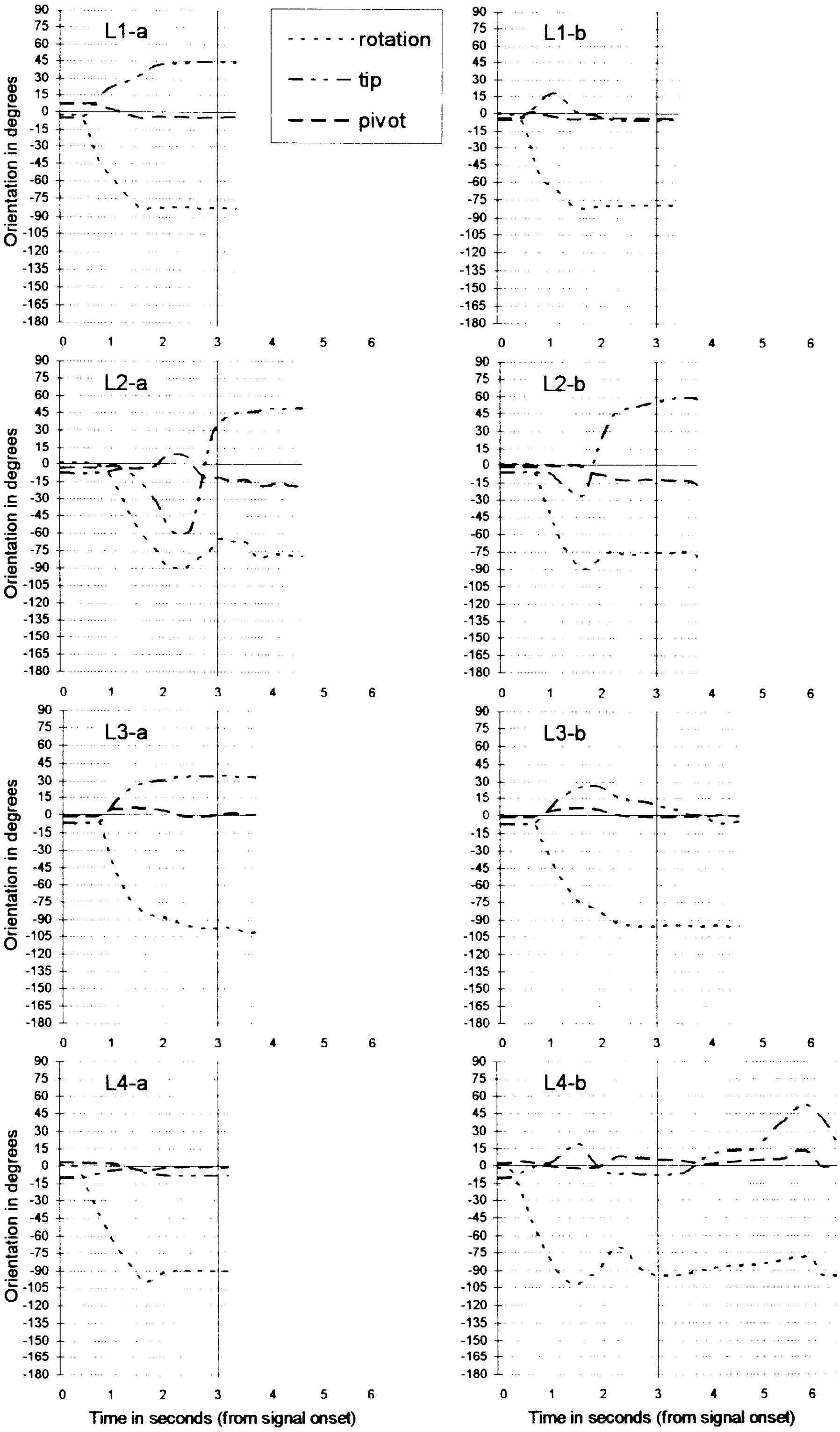

Figure 14, part 1. Head-tracker records for the four listeners, L1, L2, L3 and L4, for each of the two 3-s natural movement trials (-a, -b), involving the source at +60° elevation. Each graph shows rotation, tip and pivot records from signal onset to moment of response. Vertical line at 3 s indicates signal offset.

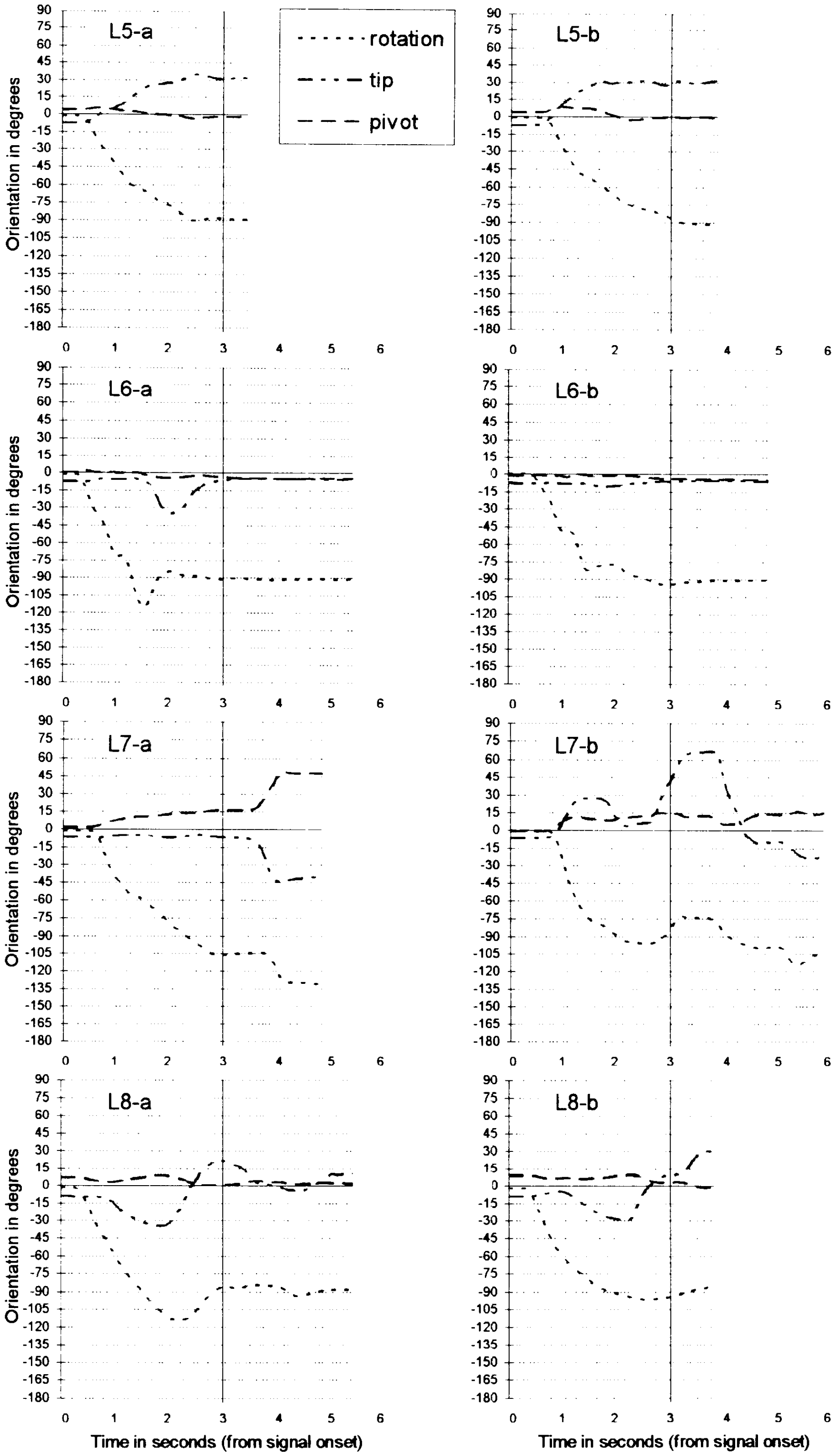

Figure 14, part 2. Head-tracker records for the four listeners, L5, L6, L7 and L8, for each of the two 3-s natural movement trials (-a, -b) involving the source at +60° elevation.

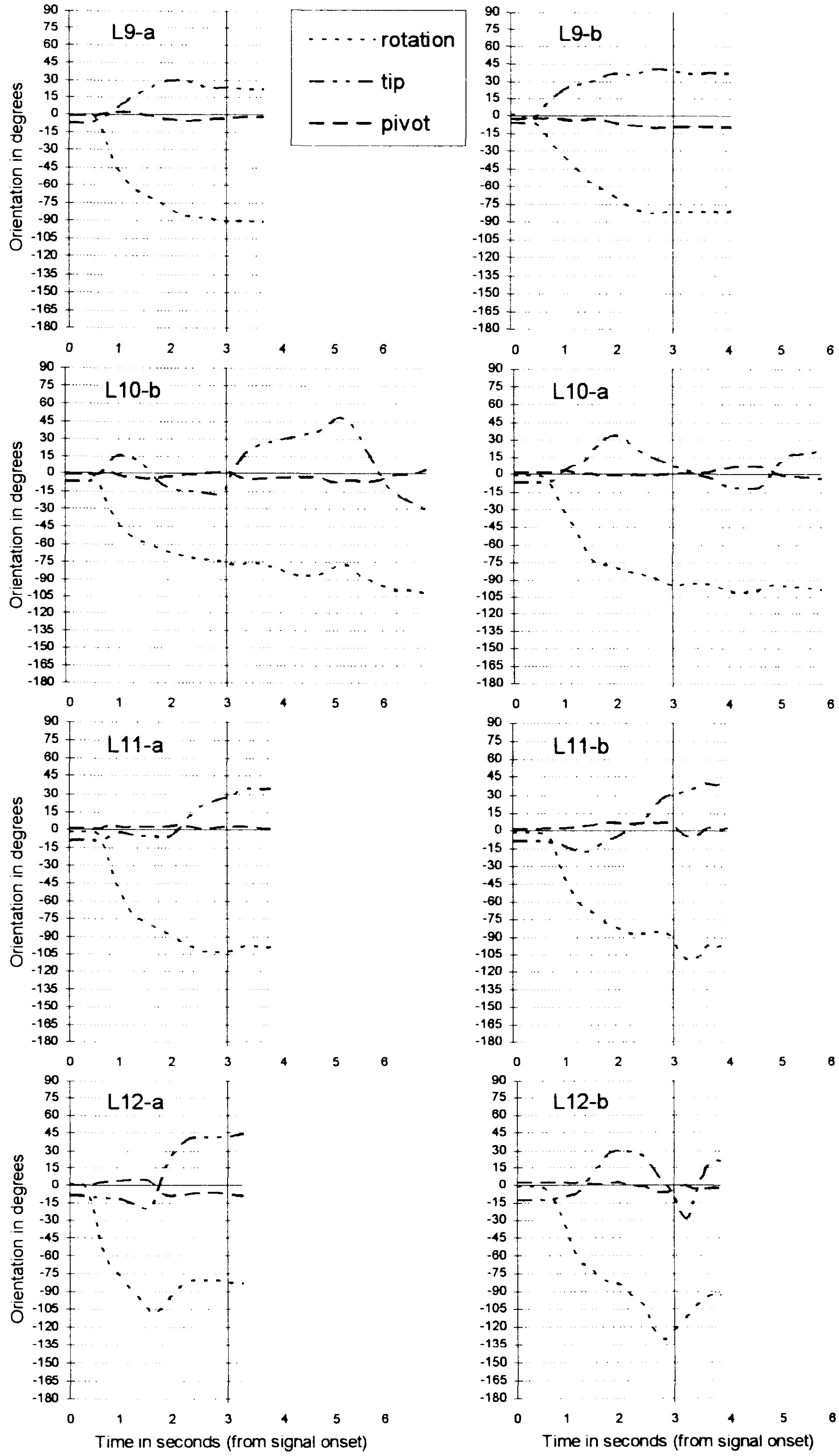

Figure 14, part 3. Head-tracker records for the four listeners, L9, L10, L11 and L12, for each of the two 3-s natural movement trials (-a, -b) involving the source at +60° elevation.

some time after rotation was initiated and is evident in Figure 14, parts 1, 2 and 3, which graph the head-tracker records, from signal onset to time of response, for all 24 trials occurring in the 3-s natural movement condition involving the loudspeaker at +60° elevation.

**Horizontal error.** Examination of horizontal error occurring in the conditions of this experiment throws further light on the matter raised by Middlebrooks and Green (1991), that head movement may have little effect except as it allows recruitment of the optimum static localization cues that become available when the source is faced. Horizontal error is the angle formed between the line passing from the centre of the listener's head to the actual source and the vertical plane passing through both the centre of the listener's head and the apparent source position. Thus, horizontal error is essentially the (double-pole) azimuth of the source with reference to the listener's orientation at the moment of responding. Horizontal error for HP sources, for each of the 3-s conditions, is plotted in Figure 15, with data involving front-back confusions and/or lack of movement in movement conditions disregarded. In the motionless listening condition, a substantial (10°-20°) horizontal error occurs for most HP sources, especially for those in the front. In other words, listeners are not able to accurately face the source if no movement occurs during signal presentation. By contrast, in both the

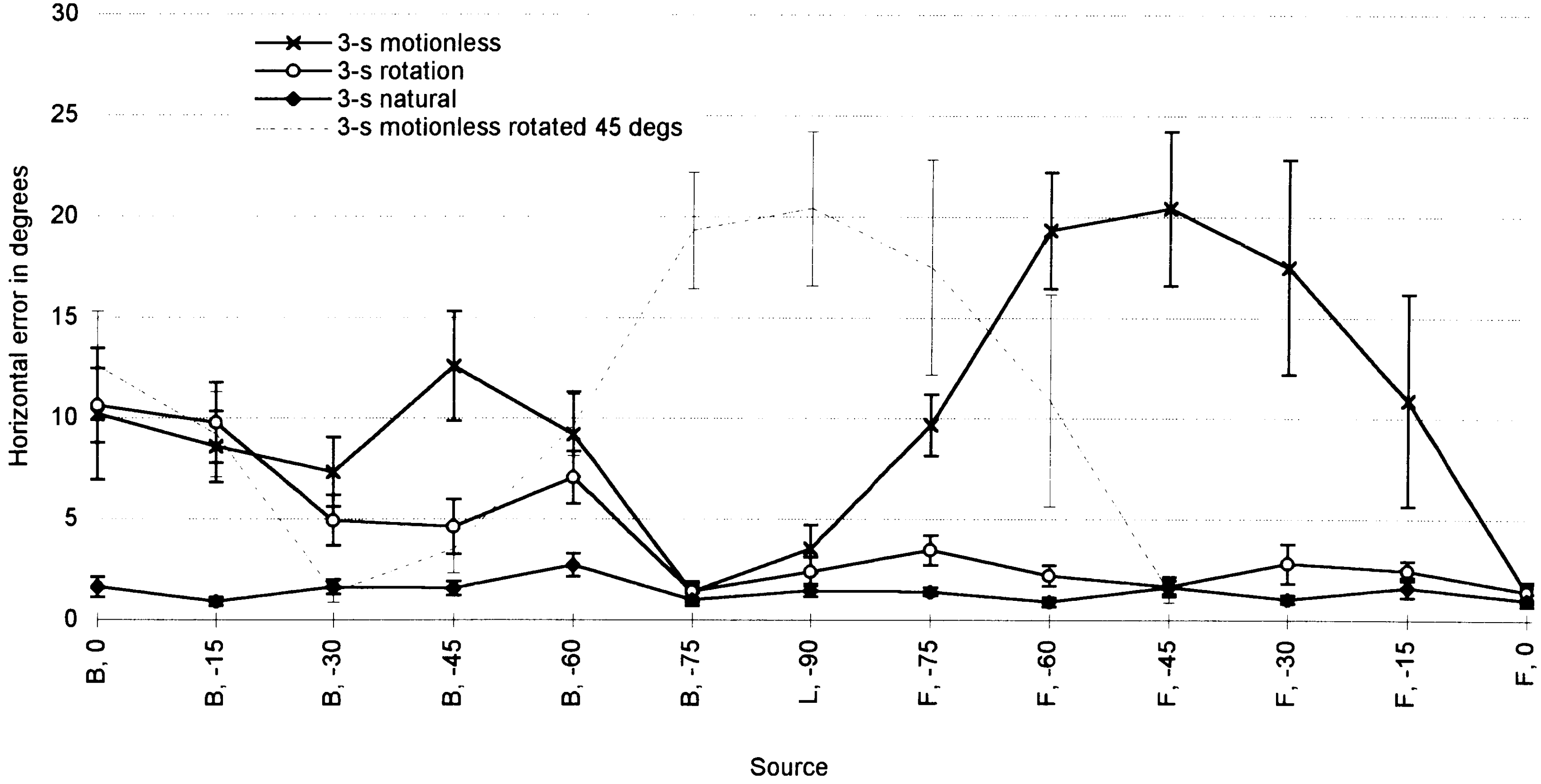

Figure 15. Horizontal error occurring for each condition with a 3-s signal. Source positions are at 0° to -75° azimuth in the back HP (B, 0 to B, -75), -90° directly left (L, -90) and -75° to 0° in the front HP. Error bars indicate standard error of the mean.

rotation and natural movement conditions, horizontal error for all sources in the front HP was always less than 4°.

Substantially less horizontal error occurred in the rotation condition than in the motionless condition for HP sources displaced up to 60° from the point faced at signal offset. In Figure 15 the motionless data are re-shown rotated through 45° (the stationary end-point of the rotation) to facilitate comparison between motionless and rotation conditions. Thus, compared with motionless listening, it is clear that rotation allows listeners to more accurately face sources positioned between 15° and 60° to the left of the point faced at signal offset. For most sources in the back HP, rotation produced similar amounts of horizontal error to motionless listening. Natural movement, on the other hand, allowed listeners to face sources in the back HP with the same high level of accuracy as it did for sources in the front HP.

Figure 16 shows the horizontal error occurring for HP sources under the different conditions with the 0.5-second signal. Again data involving front-back confusions, and non-movement within the movement conditions, are not considered. The general outcome for the 0.5-s motionless condition is similar to that of the 3-second motionless condition. Horizontal error for rotation movement with the 0.5-s signal was generally greater than for the 3-s signal, with error

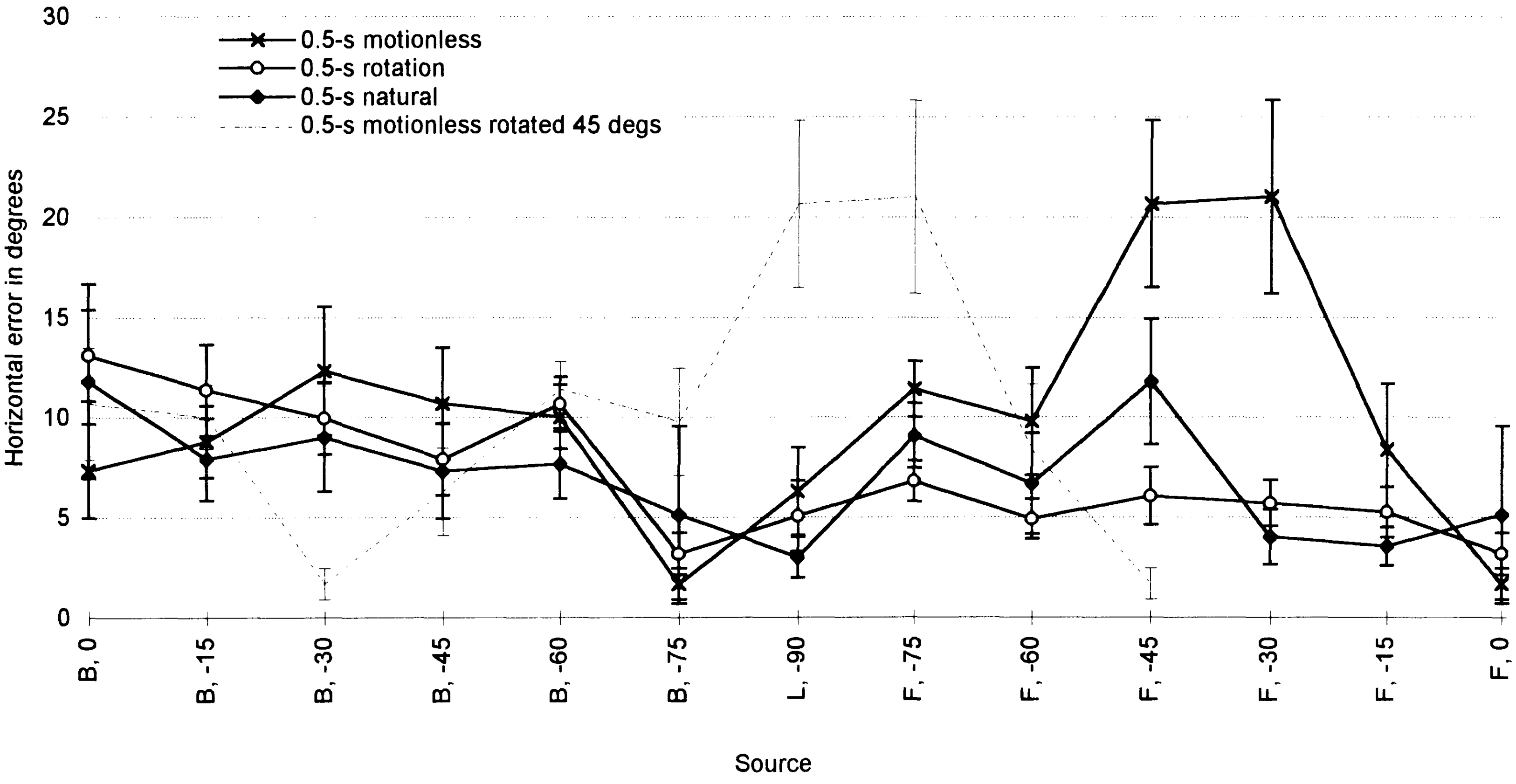

Figure 16. Horizontal error occurring for each condition with a 0.5-s signal. Source positions are at 0° to -75° in the back HP (B, 0 to B, -75), -90° directly left (L, -90) and -75° to 0° in the front HP. Error bars indicate standard error of the mean.

seldom less than 5° for positions throughout the HP. As with the 3-s signal, the pattern of responding to a 0.5-s signal reveals that listeners were able to turn and face sources in the front HP after signal offset if they had rotated, but not if they had remained motionless throughout the signal. For sources in the back HP, rotation and motionless listening produced about the same levels of horizontal error. In the natural movement condition there was a high incidence of non-movement throughout the signal, so the size of the data set for that condition is substantially reduced. With a 0.5-s signal natural listening produced substantially less error than motionless listening with front HP sources, but the greater accuracy with the 3-s signal for back HP sources, observed in natural movement compared with motionless listening, did not occur.

The 3-second rotation condition provides both facing and non-facing conditions, in the sense that performance with the loudspeaker at -45° on the HP represents a 'face-on' posture, and performance at other horizontal angles represents non-facing. Similarly, the motionless condition represents face-on to the loudspeaker at 0°. Horizontal error was very slightly lower for the source at -45° in the rotation condition than at other front HP positions, and substantially lower at 0° in the motionless conditions than at other positions. Along with the lessened accuracy for back HP sources in the 0.5-second natural condition, the data indicate that the 'face-on' argument can also be supported.

## 3.1.3. Discussion and Conclusions

Absolute accuracy. The overall result shows clearly that, with a signal of three seconds duration, both natural head movement and a single 45° rotation provide almost identical proficiency as regards the magnitude of angular deviation of responses from source positions. Whereas the 3-second natural condition may be one that enables a listener to initiate movement so as then to rely on static cues, the rotation condition limits the system, in most cases, to a set of changing interaural relationships. Wallach's (1939; 1940) claim that such a changing dynamic may be an information source gains support from this general result. The outcome of movement versus non-movement with a short-duration signal includes the complicating factor that little movement was observed in the 0.5-second natural condition. That last point is discussed later.

Front-back errors. The experiment shows very clearly that head movement contributes significantly to localization of sources in terms of resolving front-back ambiguity, even when the duration of the signal is quite short. As noted in chapter 2, a listener needs time to react to the onset of a noise burst so that the head will not rotate for the full duration of the signal. Analysis of records for the 0.5-second rotation condition revealed that the mean duration under which head rotation and noise burst occurred simultaneously was 290 ms. This suggests that the auditory system is indeed sensitive to dynamic spatial cues brought about by head rotation. That conclusion is also supported by the finding, in the natural movement condition, that front-back errors did not occur with a 0.5-second signal if a movement through as little as 5° was accomplished before signal offset. Further support for a dynamic cue to front-back discrimination is evident in the lack of front-to-back errors occurring in the 3-s rotation condition in which the angle of rotation was limited to 45°. Static interaural cues obtained at either the initial or final head position could not have prevented front-to-back errors from occurring. Such errors did occur in the motionless conditions. Again, Wallach's claim that the interaction between head rotation and source position provides information to resolve front-back ambiguity is supported by these outcomes.

There were occasional front-back errors in the region ±30° from the interaural axis. Inspection of the distribution of front-back errors showed a peak for their occurrence in that region under all conditions. Given the reduced acuity for spatial discrimination in the region around the interaural axis (Mills, 1972) some errors in that area, which were counted as 'front-back', may be better seen as instances of localization 'blur'.

Elevation. Rotation and natural movement produced significantly greater accuracy in elevation judgements than motionless listening. According to Wallach's theory mere rotation of the head would be unable to resolve ambiguity existing between elevations above the HP and those below. Although apparent elevation data suggest listeners were generally able to distinguish sources above the HP from those below, it is possible that the use of a complete spherical screen, in the present experiment, created greater up-down confusion. Some evidence for up-down confusion is associated with the source positioned directly beneath, which was generally perceived at elevations above apparent HP positions when movement was involved. Indeed, compared with the outcomes of experiments described in chapter 2, elevation judgements under natural and 0.5-s rotation movement conditions appear to be somewhat less accurate. It is reasonable to conclude that this reduction in accuracy may relate to the increase in response options available in the present experiment. But even in motionless conditions, listeners were generally able to discriminate appropriately between sources above the HP and those below. Shoulder and torso reflections may help account for this discrimination and therefore may be implicated in similar discriminations when movement is involved.

Not all listeners could use movement cues to distinguish the elevation of sources in the upper LVP, and the case of the signal overhead may explain why. Interaural differences for that signal remain null throughout any head rotation. Some listeners are sensitive to the unchanging geometry of interaural events in the face of changing head position and detect the signal as being above them. The situation may be one in which, for a signal of 3-seconds duration, the state of interaural differences at the start of a trial, prior to head movement, can be compared to their state following the cessation of such movement. Other listeners seem to attend only to final input conditions. For a source overhead, the final input condition is that the signal is in the (repositioned) MVP, and, in the absence of pinna cues for elevation, the signal is heard, relative to the listener's final orientation, as straight ahead or straight behind. With a 0.5-second signal, the head tracker records showed that the sound had ceased before movement was completed, hence the same comparison was unavailable. This may explain the markedly fewer signs of elevation detection with the shorter signal.

With sources in the lower LVP, listeners' judgements were fairly proficient, with rotation or natural movement, from 0° to -60° elevation (-90° to -60° azimuth). Even in motionless conditions, listeners could perceive the elevation of lower LVP sources to some extent — the marked apparent elevation of the lowermost source is commented on presently. The appearance of (slight) detection of the elevation of sources above and below the horizon in the absence of movement and pinna cues, suggests a role for the shoulders and torso in altering the spectrum of the signal (Gardner, 1973; Kuhn, 1987). The improvement with 45° rotation suggests the possibility of an interaction between such bodily spectral effects and changes in interaural differences. Such a possibility is outside the scope of Wallach's hypotheses, in which a side-to-side head tilt (pivot) would be thought necessary to resolve locations above and below the horizon.

There was little sign of greater accuracy in elevation detection, with movement as opposed to no movement, up to at least ±30° — the effect of movement becoming more clearly seen at greater distances from the horizon. According to Wallach (1939; 1940), this finding would be consistent with his theory because differences in the rate of change in azimuth angle relative to head orientation, for sources at different elevations within 30° of the HP, are minor compared to those beyond 30°. This point may also explain why Thurlow and Runge (1967) found little effect for movement over no movement, since their sources were only up to 41° above and below the horizon.

For the lowermost source, the observed 'lowest-to-highest' effect is a coherent outcome in one sense: the listener's seat acts to disperse energy from the lowermost region, hence there are no body-related cues along with the null interaural difference. Hence, further, there are no cues to anchor the sound to the lowermost point, and it is as plausibly heard overhead as below. This suggests, by contrast, that there are cues which anchor sources to locations in front of or behind the listener, since, in general, these were not heard as overhead or below, even in motionless listening. Shoulder and body reflections may offer such an anchor. Note that cluster 1 heard the lowermost source as at the horizon, and that this cluster also derived no benefit from rotation for the signal overhead; under that condition they heard the signal overhead as also being on the horizon in front of or behind them. For the sample as a whole, sources overhead were usually not heard as below. This may reflect a bias derived from conditions in the everyday environment.

Horizontal error. The fourth experiment provides evidence that, in general, head motion can contribute to localization accuracy even when the sound is not of sufficient duration to allow the listener to reorient so as to face the source, or, as in the rotation condition, the instruction acts to prevent this from occurring. The bulk of the effect is in the dispelling of front-back errors, but when trials involving front-back errors are discounted, listeners are still generally more accurate in rotating to face HP sources, at signal offset, compared with motionless listening. That said, there is also some evidence in favour of the point made by Middlebrooks and Green (1991). For instance, highest accuracy in horizontal localization is observed in the motionless condition at the place which is face-on throughout, and horizontal errors for brief signals, associated with truncated forms of natural movement and movement restricted to 45° of rotation, are greatest in spatial regions well away from the path of movement traversed while the source was active.

General. Results of the fourth experiment allow a conclusion that the major contribution to localization of rotational head motion cues is in eliminating large (front-back) errors. Furthermore, head movement can indicate the direction of sources lying directly overhead, and offers improved horizontal and vertical localization compared with motionless listening. There seems little additional benefit from natural movement over rotation. This suggests that information for elevation is the different transformations derived from changes in head orientation relative to resulting changes in azimuth angle: what may be dubbed the 'Wallach cue'. Evidence bearing that out is in the head tracker records in 3-second natural movement for some listeners. These show rotational movement is predominant for HP sources off the MVP, but, in contrast, substantial up/down nodding movement becomes superimposed upon the ongoing rotation only after some rotational movement has occurred, for LVP sources greater than 45° from the horizon (Figure 10, pages 81-83). The appreciation that the source may be elevated, and uncertainty as to the direction of the elevation, are outcomes that would inevitably arise if information about source direction was derived purely from the rotational 'Wallach cue'.

Allowing for individual variability, it was generally found that low-pass LVP sources tend to be heard at equivalent cone-of-confusion loci on the horizon in motionless listening. This is consistent with an earlier report (Perrett and Noble, 1995) that the system takes the horizon as its default plane when there is no strong information for placement outside of that.

The cluster groups reflected the different test orders to some extent: 3 of the 5 in cluster 1 had natural movement as their first condition, 3 of the 4 in cluster 2 began with no movement, and all three in cluster 3 began with rotation. It is not clear how the different orders of test would influence behaviour across conditions; counterbalancing was used precisely to control for any effects. There may be interactions between performance and orders of testing, and these interactions may be observed in different people to different degrees.

The general finding from the fourth experiment, in which responding was essentially totally free, is that head rotation, occurring at the same time as a sound lacking high-frequency energy, is likely to increase localization accuracy. Evidence for increased accuracy in orienting horizontally to face sources was obtained for both long and short duration signals. In the case of long duration signals, greater accuracy also occurs in the vertical horizontal dimension. The most robust effect of movement is undoubtedly the elimination of front-back confusion. The elimination of front-to-back errors with rotation of just 45° indicates listeners are able to utilize information obtained from at least two different orientations in resolving front-back ambiguity which in turn shows that head rotation provides dynamic cues rather than merely facilitating better use of static cues. Elimination of front-back confusion occurs even if the sound is of short duration and there was evidence that only very small changes in head orientation are needed. This last point warrants further investigation, specifically to test front-back discrimination with very small amounts of rotation.

## 3.2. EXPERIMENT 5

The apparent elimination of front-back errors observed in the fourth experiment with as little as 5° rotation merited further investigation. The fifth experiment directly investigated the effect of small amounts of head rotation (4° and 8°) on front-back discrimination.

The investigation of the limits of the spatial acuity of the auditory system, with respect to head motion and front-back discrimination, runs parallel to minimum audible movement angle (MAMA) research. Measurements of the MAMA have been obtained by gauging the smallest change in position of a horizontally moving sound source necessary for motionless listeners to detect the correct direction of movement (Harris & Sergeant, 1971; Perrott & Tucker, 1988). Detection of the direction of movement is particularly relevant to front-back discrimination. According to Wallach's theory, the change in direction of a sound source relative to the orientation of the head that occurs as a result of head movement produces a pattern of changing interaural relationships that enable listeners to distinguish between front and back.

It makes no difference to the interaural relationships if the head is moving and the source is stationary or the head is stationary and the source is moving. If the position of the source relative to the orientation of the head is the same, the interaural relationships will be identical. This suggests that detecting front-back information from cues resulting from head rotation would involve the same kind of auditory information as that used for detecting the direction of a moving source. The essential difference between tasks involves information about the head movement. The task of actively using head motion for front-back discrimination of a stationary source depends on information about the changing orientation of the head. Of course, such information is not relevant when the listener is motionless and the task is to detect the direction of a moving source.

Measurement of MAMAs has typically involved pure tone stimuli. Perrott and Tucker (1988) measured MAMAs, for a source moving horizontally within a 20° region directly in front of the listener, with source velocity ranging from 8°-128°/s, and found MAMAs were generally smaller for pure tones below 1 kHz than for those above. Chandler and Grantham (1992) obtained the same results with a task requiring listeners to gauge whether a source was moving or not. These researchers also found that MAMAs were greater at 60° azimuth than at 0° azimuth. On tests using noise stimuli centred at 3 kHz, they found that the broader the bandwidth the smaller the MAMAs. Since much of the MAMA research has involved pure tone signals, the decision was taken to include pure tones as well as low-pass noise in the fifth experiment, to allow comparisons with findings from such research.

## 3.2.1. Method

This experiment involved 12 people (6 females, 6 males), all reporting normal hearing. None had previous involvement in such experiments. They were presented with sounds from 8 loudspeakers within the horizontal plane. Four were to the front, positioned at 15° intervals from 0° through to -45° and four to the back from 0° through to -45°. The loudspeakers were concealed from the listener's view by the same spherical screen as used in the fourth experiment.

Four different signals of 2-s duration with 20-ms onset and offset times were created digitally. These were: 2-kHz low-pass noise, as used in the fourth experiment, and 5-kHz, 1.5-kHz and 0.5-kHz pure tones. The specific choice of frequency for the tone signals was based on MAA and MAMA data which show that MAAs and MAMAs are smallest around 0.5 kHz, greatest around 1.5 kHz and at intermediate levels around 5 kHz. All other aspects of signal production were identical to those of the fourth experiment.

Listeners were required to localize the sounds in each of three different movement conditions: 1) rotating the head 8° leftward on the HP, after the onset of the signal, and thereafter remaining motionless; 2) rotating the head 4° leftward on the HP, after the onset of the signal, and thereafter remaining motionless; and 3) remaining motionless throughout. Half of the listeners encountered these conditions in the listed order while the others encountered them in reverse order.

In all conditions, a dot of light was back-projected from a light emitting diode (LED) onto the cloth screen at 0° azimuth in the front horizontal plane. A further similarly produced light dot was back-projected on the HP at -8° azimuth for the 8° rotation condition or at -4° azimuth in the 4° rotation condition. As in the fourth experiment the light dots were used in combination with the laser pointer to allow listeners to monitor their head orientation and to gauge the extent of their head rotations in order to meet the requirements of each condition. Apart from the smaller amounts of head rotation, the procedures followed by listeners' were the same as those in the fourth experiment.

For each movement condition there was a block of 32 trials involving the presentation of each of the four signals from each of the eight loudspeakers. The orders of signal presentation and speaker activation were random. As in the fourth experiment, head tracking equipment was used to register head movements and listeners' responses, and closed circuit TV was used to monitor the listeners' movements. To aquaint listeners with the requirements of the experiment, eight practice trials were provided before each movement condition. These involved two presentations of each signal from a ninth loudspeaker positioned on the interaural axis, directly left of the listener. With minimal practice, listeners were observed to satisfactorily achieve all the required movements in each condition. As with the previous experiments, listeners received no feedback concerning localization accuracy during any part of the experiment.

### 3.2.2. Results

**Front-back errors.** Figure 17 reveals that front-back errors occurred in every movement condition with every type of signal. With 5-kHz and 1.5-kHz pure tones, all movement conditions produced front-back error rates of approximately 29% or greater, across all azimuth positions. When the signal contained lower frequency energy, an effect for the 8° rotation was apparent. With the 500-Hz pure tone, the 8° rotation condition produced front-back error rates of only 12.5% for the azimuth positions, 0° and -15°. At -45° and -30° the 8° rotation movement produced error rates of about 29% or more, which is the same general result

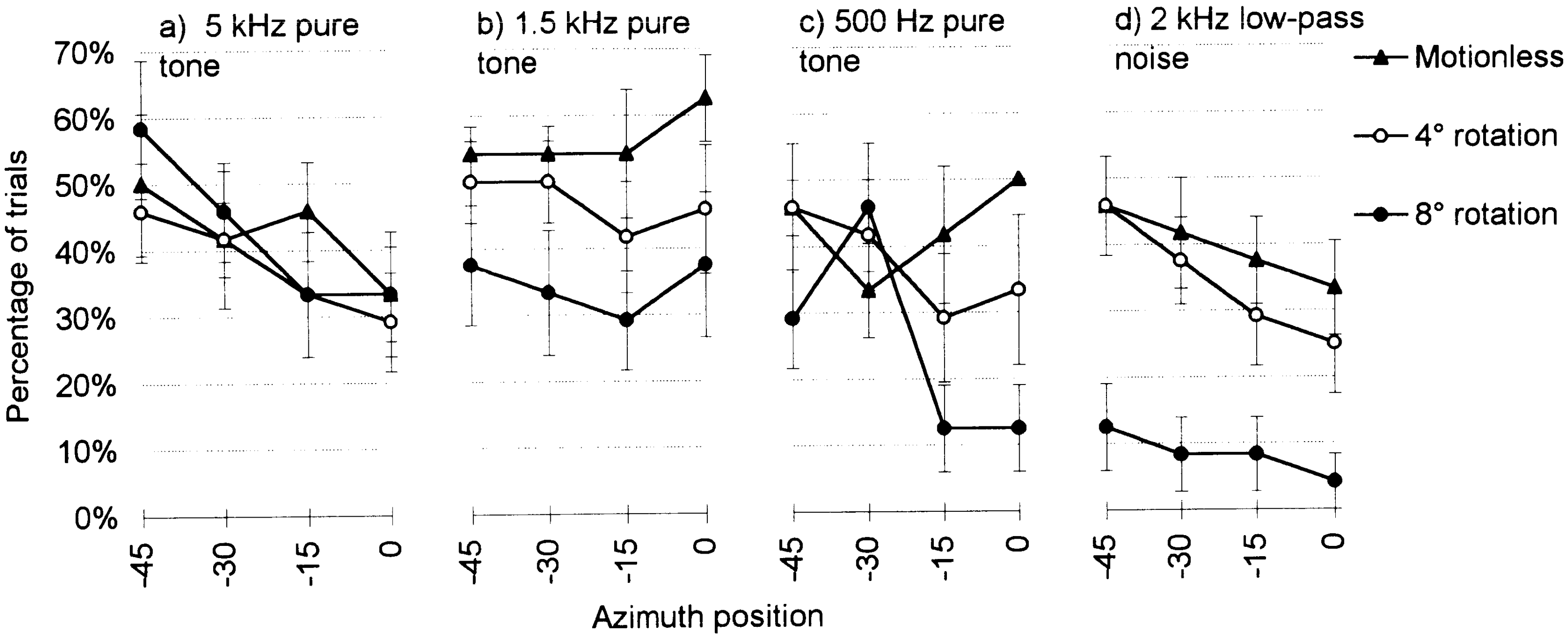

Figure 17. Front-back errors occurring for each signal type and each condition in Experiment 5. Error bars indicate standard error of the mean.

across all positions with the 4° rotation and motionless conditions. With the 2-kHz low pass noise, 8° rotation produced error rates of 12.5% or less across all positions. This is in marked contrast to outcomes from the motionless and 4° rotation conditions which produced error rates of 25% or more.

### 3.2.3. Discussion

The finding from the previous experiment, that front-back errors can be eliminated with very small amounts of head rotation is mainly supported. Although 8° rotation was unable to eliminate front-back errors completely, with the 2-kHz low-pass noise, the reduction was substantial. The results show that 8° of head rotation, and, to a lesser extent, 4° of rotation, can reduce the amount of front-back confusion only if the signal contains low-frequency energy. Head rotation of 8° with a 0.5-kHz pure tone produced a substantial reduction in front-back confusion for some sources, but not with 1.5-kHz and 5-kHz pure tones. Thus the auditory system is more sensitive to head motion cues when the signal contains low-frequency information than if it does not. This outcome is similar to that obtained from MAMA research with pure tone signals, which shows listeners are less sensitive to movement if the frequency of the signal is above about 1 kHz (Perrott and Tucker, 1988). Since, in normal circumstances, the auditory system is able to detect interaural time/phase differences and not interaural level differences in the low-frequency acoustic energy, and since a pure tone does not contain spectral information, the effect of rotation on front-back discrimination for a 0.5-kHz pure tone suggests head motion cues for front-back discrimination can be derived solely from interaural time/phase relationships.

A generally stronger effect for the 2-kHz low-pass noise than for the pure tones suggests that the rotation cue is more effective when the signal is complex. This finding is somewhat in line with those of Chandler and Grantham (1992) who showed that MAMAs are smaller with increasing bandwidth. Since the low-pass signal did not contain frequencies above 4 kHz, an effect of the pinnae is not implicated.

Experiment 5 provides evidence that the effect of head rotation on resolving front-back ambiguity is strongest for sources in or near the MVP. With the 500-Hz pure tone, 8° of rotation produced a substantial reduction in front-back confusion for the sources in or near the MVP but not elsewhere. Such a result is further evidence in line with that showing MAMAs increase with increasing azimuth (Chandler and Grantham, 1992).

One factor, important in MAMA measurement, but not controlled for in this experiment was velocity of movement. It was evident, as might be expected, that there were considerable differences between listeners in velocity of head rotation. Even within the same listener, head speed varied from trial to trial. Another factor not controlled for was the smoothness of the rotational sweep. To achieve the small amounts of rotation, that this experiment required, listeners movements were often jerky, with velocity of movement accelerating and decelerating a number of times throughout each rotation. Such variations in movement may be the reason why 8° of rotation did not completely eliminate front-back confusion in the present experiment.

Wallach provided evidence that the visual and the vestibular systems provide information about head movement. Small amounts of movement such as 4° and 8° of head rotation are not typical for natural movement, as the head tracker records for Experiment 4 indicate. Since the head rotation cue depends on information about change in head orientation, the small amounts of movement involved in the 4° and 8° rotation conditions may not have been fast or sudden enough to produce strong indications of the direction of movement, whereas the 45° rotation and the natural conditions in Experiment 4 were.

The difficulty of moving such small amounts as 4° and 8° may also have affected the results. Listeners may have had to focus too much attention on the task of moving small amounts that they failed to direct enough attention to the localization task itself. The outcomes of this and the previous experiment reported, nonetheless clearly indicate that listeners can make use of very small amounts of head rotation in order to discriminate between front and back.