Dominance of head-motion-coupled directional cues over other cues during walking depends upon source spectrum

William Martens*, Shuichi Sakamoto, Luis Miranda and Densil Cabrera

*Corresponding author's address: Faculty of Architecture Design and Planning, The University of Sydney, Sydney, 2006, NSW, Australia, william.martens@sydney.edu.au

Listeners who walk past a continuously presented speech sound source emanating from a fixed spatial position will typically experience veridical perception of source location. If, however, walking listeners are fitted with binaural hearing instruments that allow for the signals reaching their ears to be interchanged, left for right and right for left, the sound source is typically reported to be located in a spatial region that is reversed with respect to all three spatial axes: left for right, front for back, and above for below. This result has been taken as evidence for the relative dominance of dynamic interaural directional cues over the spectral directional cues associated with the pinnae of each listener, which should support veridical perception. In order to investigate the relative importance of the spectral energy distribution of the source on the illusory reversals of source location, bursts of broadband noise were presented rather than continuous speech. Under these circumstances, with greater energy in higher frequency bands, the reversals did not readily occur. Therefore, it has been concluded that head-motion-coupled directional cues are likely to dominate spectral cues associated with the filtering effects of the listener's pinnae only for sources containing greater energy at lower frequencies.

Published by the Acoustical Society of America through the American Institute of Physics
INTRODUCTION

The topic of this paper is the auditory perception of sound sources that seem to move through space for a walking listener, but are perceived as correctly stationary when the listener holds still. Such auditory illusions of source motion that depend upon listener motion were reported many years ago (e.g., Young, 1928; Willey, et al., 1937). Those publications reported anecdotal evidence on the nature of the hemifield reversals in source location, for example, between angles of incidence in the frontward and rearward hemifields. More recent experimental results using less mobile listeners, with head rotation but no translation through space (e.g., Brimijoin and Akeroyd, 2012; Morikawa, Toyoda, and Hirahara, 2011; Macpherson, 2009), provide a clearer picture of when the classically reported front/rear reversals could be expected. In particular, these provide an explanation regarding stimulus conditions under which dynamic interaural cues available to moving listeners might come to dominate spectral directional cues that are due to the filtering effects of the pinna. Unlike a great deal of research the relative dominance of dynamic interaural directional cues, however, only a few recent studies (Martens and Kim, 2009; Martens, Cabrera, and Kim, 2009) were concerned with the illusory motion of a speech sound source presented continuously over seconds (rather than milliseconds), under conditions in which listeners were instructed to walk through a fairly normal reverberant space while attempting to localize that sound source. These studies investigated the auditory illusion of source motion that occurs when walking listeners wore a binaural hearing instrument that effectively interchanged their ears, left for right (i.e., allowing the sound arriving at the blocked entrance of the left ear canal to be captured and reproduced via an acoustically isolated earphone inserted into the right ear, and vice versa). In this ‘Interchanged’ listening mode, speech sound sources that were actually located on the listener’s left side were naturally perceived to be located on the right, and vice versa. However, other reversals occurred, despite the apparently adequate preservation of pinna-based cues to front/rear and above/below source locations.

Because the binaural hearing instrument employed miniature microphones positioned at the blocked entrance to the listener’s ear canals, spectral cues to source azimuth and elevation that are associated with the filtering effects of the head and pinna could be adequately maintained. Therefore, when the ear signals were switched to the ‘Interchanged’ listening mode, a relatively normal listening experience resulted when the listener held still, despite the left-right reversal of source locations. What is most interesting about the comparison between this and the ‘Normal’ listening mode is what happened when listeners were instructed to walk toward a talker who is producing continuous speech. Though in this ‘Interchanged’ listening mode the stationary physical source of a speech sound was in clear view in front of the listeners, the auditory image of the source was invariably heard to be approaching them from behind as they approached the talker. Of course, if listeners walk straight along a path that keeps the talker on their left hand side, they will eventually reach a point at which the auditory image would be directly on their right, opposite the physical source on their left. At this point, if listeners continue along the path, the talker’s speech sound would pass them on their right-hand side, and this disembodied speech sound would continue along a path in front of the listeners. Since no physical source of the speech sound is visible, the talker then becomes an invisible ‘Phantom Walker’ moving away from the listener at twice the rate at which they are walking. Likewise, a listener who walks toward the stationary talker will experience the voice of the ‘Phantom Walker’ approaching from behind at twice the rate of closure, i.e., at twice the rate at which the listener is walking and closing the gap between the actual talker and listener.

Whereas this front/rear reversal in source locations can be explained in terms of the head turning that occurs during walking, which is not surprising as this explanation has been popular at least since it was proposed to explain Wallach’s (1939) classic results, the above/below reversals in source locations observed by Martens, Cabrera, and Kim (2011) are not so familiar nor so often reported. Although Willey, et al. (1937) reported that one of his ‘ear-interchanged’ listeners experienced the sound of a plane (that was passing overhead) as coming from the ground, they provided no explanation in terms of dynamic interaural cues related to head movements in their report. Martens, Cabrera, and Kim (2011) provide evidence from head-tracking data that the roll of the head (from side to side) during walking is the likely source of dynamic interaural cues that explain the above/below reversals. This paper will review briefly some of the results reported in Martens, Cabrera, and Kim (2011), and will extend the discussion of those effects to include the influence of source spectra on the relative dominance of dynamic interaural cues over the spectral directional cues that are due to the filtering effects of the pinna. New data will be presented from a similar study of changes in source elevation, but in a listening environment containing many loudspeakers, from which pulsed white noise was presented rather than the speech sounds that were presented in the study by Martens, Cabrera, and Kim (2011).
BACKGROUND

Although it is well established that the changes in interaural cues that occur during active localization normally aid listeners in directionally resolving environmental sound sources (see, for example, Loomis, et al, 1990), there has been little investigation the relative importance of pinna-based (spectral) cues versus interaural (e.g., temporal) cues in actively resolving source elevation angles (cf. Perrett and Noble, 1997a). More experimental evidence is available for the power of changes in interaural time and intensity differences that occur with head turning, especially with regard to the resolution of front-back ambiguity in spatial hearing by listener movement (Wightman and Kistler, 1999). Thurlow, Mangels, and Runge (1967) investigated the types and combinations of head movements that listeners typically exhibit when engaged in a free-field localization task, but did not include any investigation of the head movements that listeners make during active localization while walking towards or away from a sound source (as was done in the study by Loomis, et al. (1990). In fact, research in other sensory modalities, such as vision (Durgin, 2009), has produced results that suggest that changes occurring during walking are particularly interesting, and suggest that the study of sound localization during walking might provide a particularly rich area of inquiry. A more thorough review of the literature is beyond the scope of this paper, but it is proposed that sufficient evidence exists to support the theory that various types of head rotations occurring while listeners walk through a reverberant environment may play a dominant role in resolving the azimuth and elevation of continuously presented sound sources (Martens, Cabrera, and Kim, 2011). Before embarking on a brief summary of recent studies on this topic, a review will be presented on some earlier studies investigating the role of head movements in sound localization.

Observation of Head Movements and their Role in Sound Localization

In what is thought to be the earliest controlled study on the role of various types of head movements in sound localization, Thurlow and Runge (1967) focused upon the influence of involuntary head movements: In effect, they studied localization performance by manually inducing the head movements of their listeners rather than allowing them to generate their own head movements voluntarily. They examined errors in both azimuth and elevation judgments for a number of types of angular variation in the listener’s head, but included no translational movements (i.e., linear displacements). The three types of head angle variation that they studied are depicted graphically in FIGURE 1., and include variation in yaw, roll, and pitch. Without delineating the specifics of their experiments, their general results may be summarized as follows: Relative to a condition in which no head movement was allowed, yaw variation (aka rotation of the head) reduced errors in azimuth judgment as would be expected. However, such yaw variation (depicted in panel a of FIGURE 1) did not significantly reduce errors in elevation judgments. If, on the other hand, the listener’s head was rolled from side to side while listening (as depicted in panel b), so that first one ear was dropped closer to the shoulder on the same side, and then the other was dropped towards its shoulder, elevation error was reduced and azimuth error was not. When the head was pitched forward and back (as depicted in panel c, and characterized as looking up then looking down, as when indicating a “yes” response), neither error rate was reduced significantly. These results are consistent with how Martens, Cabrera, and Kim (2011) explained the elevation reversals observed when listeners experience the ‘Phantom Walker Illusion’ (as will be discussed more thoroughly below).

In order to determine how frequently human listeners use different types of head movements as an aid in their active determination of the location of a sound source, Thurlow, Mangels, and Runge (1967) observed over 50 subjects during a free-field localization task. First, they noted that most subjects always included some amount of horizontal head rotation (i.e., yaw variation) in their exploration of the sound stimulus. Interestingly, they observed that their listeners rotated their heads 42° on the average when the sound source was located at a high elevation, but only rotated their heads an average of 29° when the source was at a lower elevation. In fact, it was most common to find listeners rotating their heads while also pitching their heads (which they termed tipping). The way in which they classified observed head movements used a criterion of 3° for each of the three types of variation in head orientation. Their results can be summarized as follows: The type of head movement observed most often was a combination of rotation and pitching, without substantial rolling of the head (which they termed pivoting). The second most common type of head movement observed was rotation without pitching or rolling, while the third most common was a combination of rotation, pitching, and rolling. Fourth was rotation with rolling. The relatively high frequency of this fourth type of movement was reported as somewhat surprising by Thurlow, Mangels, and Runge.
(1967) because they observed that head rotation and roll produced confounding lateral shifts of the auditory image that should not disambiguate direction. That is, because rotation-induced binaural changes most strongly cue front/rear distinctions, and roll-induced binaural changes most strongly cue above/below distinctions, the interpretation of the binaural changes accompanying combined rotation and rolling movements were regarded by Thurlow Mangels, and Runge (1967) as providing only uncertain information.

a) **Yaw angle variation:**  
Angular acceleration of a listener’s head around the pictured vertical axis, as when one is indicating a “no” response (i.e., turning the head to the left and right).

b) **Roll angle variation:**  
Angular acceleration of a listener’s head around an axis extending from front to rear (axis not pictured), so that the head rolls from side to side.

c) **Pitch angle variation:**  
Angular acceleration of a listener’s head around the pictured horizontal axis extending from left to right, as when one is indicating a “yes” response.

**FIGURE 1.** Diagrams showing three types of angular acceleration (change in orientation) of a listener’s head: 1a) head turning (yaw angle variation), 1b) head pivoting (roll angle variation), and 1c) head tilting (pitch angle variation). To best understand these three types of angular acceleration, one can imagine axes about which rotations occur given a listener located at the origin of a Cartesian coordinate system. When a listener engages in head turning, rotation occurs around a vertical axis (pictured as a vertical dashed line in panel 1a); when engaged in head pivoting, rotation occurs around an axis extending from front to rear (not pictured in panel 1b, as it runs perpendicular to the plane of the page); and when engaged in head tilting, rotation occurs around an axis extending from left to right (pictured as a horizontal dashed line in panel 1c).
Studies of an Auditory Illusion of Source Motion During Walking

Although there are several notable and important examples of research involving auditory illusion of source motion (e.g., Wallach, 1939), there is very little literature on the conditions under which such illusions may occur for listeners walking past a presented sound source. Nonetheless, it is clear from the combined results of studies by Thurlow and Runge (1967) and Thurlow, Mangels, and Runge (1967) that either voluntary or involuntary head movements may provide information that can aid listeners in localizing sound sources in the three-dimensional space surrounding them. So whether listeners are consciously aware of the various types of head movements in which they are engaged while walking or not, there in no reason to suppose that they do not have access to the relevant multimodal information. Indeed, subsequent to his more theoretical 1939 paper, Wallach (1940) reported on the role of head movements and vestibular and visual cues in sound localization, and reached a similar conclusion. But what is relevant to the current study is the extent to which the three identified head rotations might be involved in active sound localization while walking. For a more thorough review of this topic, the reader is referred to Martens, Cabrera, and Kim (2011). Suffice it to say that variation in both yaw and roll angles of the head is increased when a stationary listener begins to walk, but it is the variation in head roll angle that sees the greatest increase. So, assuming that listeners will use the available information to resolve conflicts in localization cues, a series of studies were undertaken to investigate this potential dominance of head-motion-coupled directional cues over spectral cues associated with the filtering effects of the listener’s pinna.

Martens and Kim (2009) made the earliest report on this topic, summarizing the results of a simple pilot study in which five walking listeners completed a source elevation judgment task in a single experimental condition. The listeners were fitted with a binaural hearing instrument that allowed for the signals reaching their ears to be interchanged, and no normal listening condition was included in the study. As explained in this paper’s introduction section, the binaural hearing instrument was configured in such a way as to preserve the spectral cues associated with each listener’s own pinna, and at the same time allowed a speech sound source arriving at the blocked entrance of the left ear canal to be captured so that it could be reproduced via an acoustically isolated earphone inserted into the right ear, and vice versa. Listeners were required to report which of five loudspeakers seemed to be producing a continuous speech sound as they walked past a vertical array with two of the five loudspeakers above ear level, and two below. As the observed loudspeaker elevation reports were inversely related to the actual elevation angles of the loudspeakers, it was concluded that interchanging the listener’s ears caused them to rely more upon the reversed interaural cues than they relied on the usual, and supposedly adequately maintained, spectral cues to elevation. However, this early study did not test whether elevation judgments produced veridical elevation percepts when listeners used the binaural hearing instrument in the normal (non-interchanged) listening mode. Therefore, Martens, Cabrera, and Kim (2009) executed a more comprehensive study in which listeners were tested under a variety of conditions.

In order to determine if walking was required for the listener to experience a reversal of elevation percepts when wearing the binaural hearing instrument that effectively interchanged the listener’s ears, Martens, Cabrera, and Kim (2009) had listeners give elevation reports under four conditions defined by a two-by-two factorial combination of treatments: Each listener had the binaural signals delivered by the hearing instrument in either the normal or the interchanged mode, and furthermore each listener was either walking or standing still in a given run of experimental trials. For sake of controlled comparisons, the ear signals could be quickly switched between the ‘Interchanged’ and the ‘Normal’ listening mode, so that dynamic interaural and spectral directional cues would give either consistent or contradictory information regarding source locations. One further difference between this study and the pilot study (Martens and Kim, 2009) was that listeners were required only to report whether the source was located above or below ear level. As in the pilot study, a continuous speech sound was presented from one of five loudspeakers as listeners walked past the vertical array, with two loudspeakers were positioned above ear level, and two below. In the case of sound stimuli reproduced via the fifth, ear-level loudspeaker, listeners still were forced to report whether the source seemed to be arriving from a loudspeaker angle that was either above or below ear level. Therefore, the observed elevation judgments, which are shown for all four experimental conditions in FIGURE 2, are plotted in terms of the proportion of ‘Above’ responses (summing over all five listeners in two different listening modes and under two different listener movement conditions). The results of this study (which are explained in greater detail in the figure caption) established that the experimental interchanging of ear signals did not cause a reversal of elevation judgments when listeners were asked to stand still. Yet when listeners with interchanged ear signals walked past a continuously presented speech sound source, it was heard to be located in a spatial region that was reversed with respect to all three spatial axes: left for right, front for back, and above for below.
FIGURE 2. Proportion of ‘Above’ responses resulting from summing the elevation discrimination data obtained from five listeners in two different listening modes and under two different listener movement conditions. The panel on the left shows their forced choice discrimination performance in the ‘Standing Still’ movement condition, and the panel on the right shows performance in the ‘Walking’ movement condition. In both panels, circular symbols are used to plot response proportions as a function of the elevation of the sound source when listeners were in the ‘Normal’ listening mode (in which the left earphone emitted the left earmic signal, and the right earphone the right earmic signal); the ‘x’ plotting symbol in both panels was used to display performance when listeners were in the ‘Interchanged’ listening mode (in which the left earphone emitted the right earmic signal, and vice versa). The solid red curves show the logit functions fit to the ‘Above’ proportions obtained in the ‘Normal’ listening mode, while the dashed red curves show the fit to the ‘Interchanged’ listening mode data.

As in the pilot study, though the stationary physical source of the sound was in clear view as listeners walked toward it, the sound was invariably heard to be approaching them from behind, and the voice of this illusory ‘Phantom Walker’ overtook listeners as they passed by the physically stationary source. Although this is an interesting illusion, it presents a problem for unbiased observation of listener performance, and this is a real drawback in the designed experiment that was recognized by the experimenters. Since the single vertical loudspeaker array was always located on only one side of the listener, experienced listeners came to recognize whether they were listening in the ‘Normal’ versus the ‘Interchanged’ mode by virtue of the difference that occurs when they walked past a stationary sound source on the right: In the ‘Normal’ mode the auditory image associated with the source is also on the right, while in the ‘Interchanged’ mode the auditory image is reversed, and appears on the left side of the listener. Therefore, the experiment could not be considered a double-blind test of the hypothesis that elevation percepts would be reversed in the ‘Interchanged’ mode. In order to keep the subject less aware of the listening mode, identical vertical arrays of five loudspeakers would need to be placed on both the left and right side of the listener. However, this only removed the telltale clues for the stationary listener, since the walking listener would hear the motion of the ‘Phantom Walker’ source in the ‘Interchanged’ mode only. In order to keep the listener as naïve regarding experimental manipulations as possible, it was decided to repeat the study in a unique room in a Tohoku University laboratory, which is described in Okamoto, Zheng, Iwaya, and Suzuki (2010). This room contained 157 loudspeakers roughly evenly distributed along all four walls and the ceiling of the space. Thus, as far as listeners knew, the auditory imagery they experienced on a given trial could have been produced by any combination of the 157 loudspeakers (although only one loudspeaker actually produced the sound sources presented in this study). Another difference in this subsequent study, which is reported in this paper for the first time, was that results given a new sound source, a pulsed white noise, could be compared to results for a continuous speech signal.
METHODS

Under four distinct conditions, a continuous sound stimulus was reproduced via one of five loudspeakers out of the 157 loudspeakers that were distributed evenly about the room through which listeners could walk during listening trials. The set of five employed loudspeakers were arranged vertically such that one of the loudspeakers was located at the same height as the subject’s ears. Two of the loudspeakers were located above ear level, and two below, with spacing at roughly 20-degree intervals relatively to the adjacent (i.e., stationary) listening position, so that the overall vertical extent of the array ranged roughly from -40° to +40° in elevation (this was the extent when the array was positioned directly to the listener’s side, at 90° azimuth). Of course, the effective vertical extent for the walking listener was only momentarily covering this range as the listener walked past the active loudspeakers. The task for the listeners was to indicate whether a continuously presented sound stimulus was arriving from above or below ear level. In the ‘Standing Still’ condition, listeners were asked to keep their heads as still as possible, and to make an elevation choice as soon after the onset of the sound source as possible. In the ‘Walking’ condition, listeners were asked to wait until the playback of the stimulus had begun before beginning to walk toward the vertical array of loudspeakers, and to wait until after they had walked past the vertical array of loudspeakers before making an elevation choice.

The task required a two-alternative forced choice (2AFC) response as to whether the source sounded as if it above or below the listener’s ear level, with no option for expressing uncertainty or an intermediate ‘ear level’ direction. Ten listeners completed 20 forced choice trials in each of the two movement conditions (‘Standing Still’ vs. ‘Walking’). Within a block of trials, the connection between earmics and insert earphones was switched randomly via a relay under computer control, so the binaural hearing instrument delivered signals on each successive trial in either listening mode (‘Interchanged’ vs. ‘Normal’). In the one half of the experimental trials, the stimulus was a white noise pulse with a width of 20 ms per pulse, repeated at a rate of 10 Hz. In the subsequent second half of the experimental trials, the stimulus was a continuous speech sound (taken from Diana Deutsch’s 2003 “Speech Illusions” CD), which on average had considerably greater low-frequency content than did the broadband noise.

RESULTS AND DISCUSSION

The results obtained when pulsed white noise signals were presented for elevation judgments are shown in FIGURE 3. The proportions plotted in the left panel of the figure illustrate that it was not difficult for at least one of the 10 listeners tested to judge whether the stimulus came from above versus below ear level in the ‘Normal’ listening mode, regardless of the movement condition (‘Standing Still’ vs. ‘Walking’). These plots show the observed proportions of ‘Above’ responses when sounding loudspeakers were actually located above ear level under four conditions (see caption). In the ‘Normal’ listening mode, sources reproduced by elevated loudspeakers were almost always identified correctly (as arriving from above ear level) by this one listener. In contrast, the proportions plotted for this one listener in the ‘Interchanged’ listening mode were at chance when standing still, but showed significant elevation reversal when the listener walked past the sounding loudspeaker.

FIGURE 3. Proportion of ‘Above’ responses obtained from either one listener (left panel) or averaged over 10 listeners (right panel) in two different listening modes and under two different listener movement conditions. The light-gray bars plot forced choice discrimination performance in the ‘Standing Still’ movement condition, and the dark-gray bars plot performance in the ‘Walking’ movement condition. Pairs of bars on the right half of each panel show response proportions for the ‘Interchanged’ listening mode while bars on the right half of each panel show response proportions for the ‘Normal’ listening mode.
When elevation judgments are averaged over all 10 listeners, there is apparently no evidence for such a reversal in elevation when those listeners walked past the source in the ‘Interchanged’ listening mode, as shown on the right side plot in FIGURE 3. The proportions plotted in the right panel of the figure illustrate that it generally was quite difficult for the group of 10 listeners tested to judge whether the stimulus came from above versus below ear level even in the ‘Normal’ listening mode, regardless of the movement condition (‘Standing Still’ vs. ‘Walking’). The analysis of these results does not allow a strong conclusion to be drawn about the relative strength of dynamic interaural cues versus the spectral directional cues that are due to the filtering effects of the pinna. What the results seem to show is that listeners presented with broadband noise pulses experience unstable direction percepts, in a manner suggesting that they could attend either to spectral cues or dynamic interaural cues for these sources. The results obtained when speech signals were presented for elevation judgment are not plotted here, but are generally consistent with the results that were shown in FIGURE 2 (those from a previous study). That is, several subjects showed the elevation reversals that are expected given that they experience the ‘Phantom Walker’ illusion. Note, however, that some subjects listening to speech sounds showed inconsistent results, and seemed to shift their attention between spectral cues and interaural cues, perhaps due to previous exposure to the noise stimuli. Suffice it to say, however, that there was an indication that more subjects presented with speech sounds more often relied upon dynamic interaural cues in making elevation judgments when they were in the ‘Interchanged’ listening mode.

The proportions plotted in the left panel of FIGURE 2 illustrate that it was not difficult for listeners to judge whether the presented speech sound sources were located above or below ear level when those listeners were standing still, regardless of whether or not the signals delivered to their insert earphones were interchanged left for right with the signals captured at their ears. But what can explain the reversal in elevation judgments that occurs between ‘Normal’ and the ‘Interchanged’ listening modes for listeners who walk past the target sound source whose elevation is in question? It cannot be the head turning that was implicated in explaining front/back reversals, because the shifts in interaural differences that result from left/right head turning provide ambiguous above/below cues. Rather, it must be the rolling of the listeners’ heads during walking that is providing the disambiguating cues to elevation (as revealed in Martens, Cabrera, and Kim, 2011), which would be consistent with observed dependencies in vertical plane sound localization results found for seated listeners (Perrett and Noble, 1997a). What is apparent in their results, and in those reported for the current study, is that the natural variation in head angles that occur in active localization are complex enough to allow for a variety of results to be observed across listeners as well as across blocks of trials. This is especially true for longer duration stimuli, since more than one change in angular acceleration is possible when durations exceed a second or two. Nonetheless, when the listener is not moving, and is instructed only to make simple front/back discriminations, improvements in performance are to be expected when subjects turn their heads. For example, Perrett and Noble (1967b) found significantly fewer front-back errors for long stimuli (3s) than for short ones (0.5s), and indeed, the front/back error rates they observed for listeners holding their heads motionless were reduced significantly when listeners were allowed to turn their heads.

Role of Source Spectra in Dynamic Source Localization

The ‘Phantom Walker Illusion’ that was originally presented as providing evidence for the dominance of dynamic interaural over spectral directional cues during walking seems on further investigation to be quite dependent upon source spectra. If speech is the stimulus to be localized, then the cue conflict may be resolved in favor of the dynamic interaural directional cues. However, when pulsed white noise is the source stimulus, the spectral cues seem to carry more weight. These elevation judgment results are consistent with a number of findings in studies of front/back reversals that depend upon source spectra. For example, Brimijoin and Akeroyd (2012) report that an illusion of auditory motion in the front hemifield for a stimulus presented in the rear hemifield was experienced most strongly when that source was lowpass filtered. They wrote that “signals with the most high frequency energy were often associated with an unstable location percept that flickered from front to back as self-motion cues and spectral cues for location came into conflict with one another.” Macpherson (2009) also showed a similar dependence of such illusory sound source motion on source spectra. He found that head-rotation-coupled interaural cues would stabilize a broadband virtual source in front of a listener even though Head-Related Transfer Functions (HRTFs) for a rearward direction were used in processing that source. However, the telling result, discussed in more depth in Macpherson (2011) was that high-frequency band-limited sources, under otherwise similar conditions, were heard to move rapidly through space with the listener’s head turns. In only one previous
study of virtual source localization was there an explicit test of the relative influence of head-motion-based directional cues versus the ‘pinna factor’ when the two were in conflict. Presenting relatively long-duration musical stimuli processed using HRTFs for locations that conflicted with dynamic Interaural Time Delay (ITD)-based cues, Kawaura, Suzuki, Asano, and Sone (1991) reported that the localization in headphone reproduction was dominated by the head-coupled dynamic interaural cues.

Hirahara and Morikawa (2011) also reported that the impact of dynamic binaural signals on three-dimensional sound reproduction was quite significant, but such results are not specific to virtual auditory display of sources. In a prior free-field study, Iwaya, Suzuki, and Kimura (2003) showed that the front-back error in localizing lowpass-filtered sound was quite large unless head movement was allowed, and then only for fairly long-duration stimuli. In another free-field localization study, Morikawa, Toyoda, and Hirahara (2011) reported that the impact of head movement on horizontal-plane sound localization was greatest when band-limited noise was the stimulus, showing that head motion improved performance very little in the broadband case, and that performance dropped from around 96% to less than 70% correct when listeners were instructed to hold the head still. The plausible explanation is that narrow-band sounds are associated with apparent locations more according to their frequency content than by their actual locations, as shown in Blauert (1969), and most importantly, showing this dependence on frequency content only when the listener’s head is held still (as was emphasized in Blauert, 1997).

CONCLUSION

The current experiment confirms that some listeners wearing a binaural hearing instrument to effectively interchange their ear signals will experience a reversal of source locations in auditory space on all three axes; left for right, front for back, and up for down. However, these reversals occur more often when sources to be localized have greater low-frequency content than pulsed white noise. In particular, it was the results of the controlled investigation of elevation judgments for speech sound that support the conclusion that head-motion-based directional cues during active localization are more salient than the ‘pinna factor’ when the two are in conflict.

ACKNOWLEDGMENTS

This research was supported under the Australian Research Council’s Discovery Projects funding scheme (project DP120100484). The authors thank Takuma Okamoto, Yukio Iwaya, and Yoiti Suzuki for their support of this study and their generosity in hosting the project at the Research Institute of Electrical Communication, Tohoku University, Sendai, Japan.

REFERENCES


Suzuki et al.

[40x800]World Scientific, Singapore), pp. 159-158.


