VISUAL BIASING OF AUDITORY LOCALIZATION IN AZIMUTH AND DEPTH

BRIAN T. AGGANIS
Psychology Department

JEFFREY A. MUDAY
Biology Department
Wake Forest University

JAMES A. SCHIRILLO
Psychology Department

Perceptual and Motor Skills, 2010, 111, 3, 872-892.

VISUAL BIASING OF AUDITORY LOCALIZATION IN AZIMUTH AND DEPTH[1, 2]

BRIAN T. AGGANIS
Psychology Department

JEFFREY A. MUDAY
Biology Department
Wake Forest University

JAMES A. SCHIRILLO
Psychology Department

Summary.—Correctly integrating sensory information across different modalities is a vital task, yet there are illusions which cause the incorrect localization of multisensory stimuli. A common example of these phenomena is the "ventriloquism effect." In this illusion, the localization of auditory signals is biased by the presence of visual stimuli. For instance, when a light and sound are simultaneously presented, observers may erroneously locate the sound closer to the light than its actual position. While this phenomenon has been studied extensively in azimuth at a single depth, little is known about the interactions of stimuli at different depth planes. In the current experiment, virtual acoustics and stereo-image displays were used to test the integration of visual and auditory signals across azimuth and depth. The results suggest that greater variability in the localization of sounds in depth may lead to a greater bias from visual stimuli in depth than in azimuth. These results offer interesting implications for understanding multisensory integration.

A common multisensory phenomenon is the ventriloquism effect (Thurlow & Jack, 1973; see Bertelson & de Gelder, 2004, for a review), where bias occurs while locating an auditory stimulus in the presence of a concurrent visual stimulus at a different location (Bertelson & Radeau, 1981; Wallace, Roberson, Hairston, Stein, Vaughan, & Schirillo, 2004; Warren, Welch, & McCarthy, 1981). While this effect has been extensively studied in azimuth, no consensus has been reached regarding the ventriloquism effect in depth. The following experiment expands the literature by systematically exploring multisensory integration across both azimuth and depth. In his seminal work, Gardner (1968) found that localizations of sound in depth were strongly biased by appropriate visual stimuli. Gardner placed five inline speakers at eye level, ranging in distance from 3' to 30' such that only the nearest speaker was visible to the observer. Regardless of the actual sound-source distance, observers exclusively located the sound as originating from the closest speaker, a phenomenon which Gardner termed the "proximity-image effect." Mershon, Desaulniers, Amerson, and Kiefer (1980) examined this effect under different conditions including using normal, reverberant rooms and found that bias was

[1]Address correspondence to Dr. James A. Schirillo, Department of Psychology, P.O. Box 7778 Reynolds Station, Wake Forest University, Winston-Salem, NC 27209 or e-mail (schirija@wfu.edu).
[2]The authors especially thank Andrew Hoyord (Tucker-Davis Technologies) for programming assistance, and constructive comments from two anonymous reviewers.

DOI 10.2466/22.24.27.PMS.111.6.872-892 ISSN 0031-5125

elicited by visual stimuli both proximal and distant to the actual source. This ultimately led to replacing the term "proximity-image effect" by the more encompassing term "visual capture."

Zahorik (2001) noted that Gardner (1968) used an anechoic environment, which is known to impoverish auditory distance information because it effectively removes reverberant sound energy. However, Mershon, *et al.* (1980) differed from Gardner (1968) in terms of the number of sound sources used, their configuration in space, the listener's task in the experiment, and type of source stimulus. Thus, Zahorik (2001) attempted to replicate more closely Gardner's (1968) original conditions, while also adding an auditory-only localization condition. For example, Zahorik (2001) used a small semi-reverberant room, in which the small size allowed for extensive reverberation cues. The "sound" was a 5-sec. recording of a female voice made in an anechoic chamber. In the auditory-only condition (with observers blindfolded), observers were accurate. This effect increased when the speakers were visible, thus contradicting the "proximity-image or visual capture effect." One major difference that Zahorik mentioned was use of multiple sound sources within a block of trials, thereby providing the observer with relative distance information. This helped with the auditory-only condition and may also account for why visual capture was not found. Zahorik also suggested that he provided "rich auditory distance information with multiple potential targets in space." However, since he also did not provide single-source auditory-only data, he could not speculate further why there were differences between their findings. Taken together, however, the aforementioned studies offer some evidence that visual capture exists; the previous findings are inconsistent and often contradictory.

Present auditory stimuli used virtual acoustic technology (Wightman & Kistler, 1989a). Hypothetically, sound waves recorded from a real source in an individual's ear canal contain all of the depth and localization cues normally found in the environment. Thus, playing the recording over headphones mimicked the experience of perceiving the noise in a free-field environment and provided observers the experience of the sound originating from its original location in space.

Also utilized was stereovision as a means of effectively placing objects in three-dimensional space (Wagner, 2004). By taking advantage of virtual acoustics and 3D stereo-images, the current experiment presented stimuli without the fixed speakers and LED arrays found in prior research. This eliminated obstructions in the visual and auditory fields and provided a novel method for study of multisensory bias which systematically addressed effects across both azimuth and depth dimensions.

Sound localization depends on many factors, such as frequency, intensity, direct-to-reverberant energy ratio, and binaural differences (Ash-

mead, LeRoy, & Odom, 1990; Bronkhorst & Houtgast, 1999; Schilling & Shinn-Cunningham, 2002; Zahorik, 2002b, 2002c). Even in the presence of multiple cues, sound localization is generally poor with a large amount of variability (Wenzel, Arruda, Kistler, & Wightman, 1993; Zahorik, 2002b). As individuals generally have the worst localization for sound sources in elevation or in depth (Guski, 1990; Durlach, Shinn-Cunningham, & Held, 1993), poor and variable depth localization were hypothesized to correlate with greater visual bias in depth compared to azimuth judgments. This hypothesis is consistent with a Bayesian analysis which stipulates that as the variance increases in one signal from a multisensory stimulus, bias toward the other multisensory signal will also increase (Battaglia, Jacobs, & Aslin, 2003; Alais & Burr, 2004; Rowland, Stanford, & Stein, 2007). As Battaglia, *et al.* (2003) stressed, both the mean and the variances must be statistically independent and the prior distribution of the mean must be a uniform distribution (meaning that all values are equally likely).

Method

Participants

Three male observers (ages 21, 22, and 48 years) were recruited from the Wake Forest University perception lab. They gave their informed consent in writing. Observers JS and BA were not naïve to the ventriloquism illusion. All procedures were approved by the Institutional Review Board of Wake Forest University and were performed in accordance with the ethical standards established by the 1964 Declaration of Helsinki.

Materials

Binaural-related impulse-responses (BRIRs) were computed using a loudspeaker (Infinity Reference 3002cf 3.5 Speaker System Frequency Response 85Hz–21kHz; 75 Watts; 4 Ohms SPL 92 dB) manually positioned at nine different locations in front of a CRT monitor. BRIRs were measured by placing two Sennhesier (KE4-211-2) miniature-ear electret capsule microphones enclosed by Etymotic Research ER-13R-2 ring seals in each observer's ears and recorded sounds from each of the different locations in space. These recordings were then played back as virtual sounds, meaning that one did *not* have to compute individual head-related transfer functions. Thus, BRIRs accounted for the unique transfer characteristics of the observer's pinnea, outer ear canal, head, and torso as well as the experimental room reverberations (Savioja, Huopaniemi, Lokki, & Väänänen, 1990; Vesa & Lokki, 2006; Lindau, Maempel, & Weinzierl, 2008).

Wightman and Kistler (1989a) outlined and tested a standard method for producing virtual acoustic sounds by capturing interaural time differences and interaural level differences by using two independent probe microphones placed within the ear canal and capable of recording

the separate sound signals reaching each ear. These recordings produced head-related transfer functions which take into account the unique influence on the sound waves of an individual's torso and folds of the pinnea. Wightman and Kistler (1989b) validated their method by finding a very close correspondence between free-field and virtual-source observer judgments, providing compelling evidence that virtual acoustics can produce an accurate three-dimensional listening experience. This work was replicated by Langendijk and Bronkhorst (1999) who demonstrated that localization was identical for real sound sources and their virtual counterparts (also see Wenzel, *et al.*, 1993).

While broadband "white noise" is easier to localize than pure tones in azimuth and elevation (Carlile, Leong, & Hyams, 1997), it is unclear if such improvements are present in depth perception. Given these considerations, an impulse response using a maximum-length sequence optimal for virtual acoustics (MLS; Rife & Vanderkooy, 1989) generated by Tucker-Davis Technologies software (System 3: 2 DSP Piranha Multiprocessor for 2 channel D/A A/D conversion) was utilized instead of white noise.

Sound bursts were 100 msec. in length, while recordings lasted 200 msec. to capture room reverberations after the auditory stimulus (Mershon, Ballenger, Little, McMurtry, & Buchanan, 1989). The recorded BRIRs were played back to observers over dynamic, open, diffuse-field, studio headphones from Beyerdynamic (DT 990 Pro; Nominal frequency response 5Hz–35kHz). Reference measurements were made at the location of the observer's head, using a ½-in. condenser microphone from Brüel & Kjær (omnidirectional cartridge Type 4191, 200V, free-field, 3Hz–40kHz) enhanced by a ½-in. microphone preamplifier (Type 2669-B), and measuring amplifier (Type 2609).

To produce visual stimuli, images were presented on a 20 in. flat-profile, fixed-frequency Clinton Monoray monitor (640 × 480 spatial resolution) at 200 Hz to provide flicker-free stereo-fusion (100 Hz to each eye on alternate frames). The yellow stereo-images were circular, with Gaussian blurred edge, 1° visual angle light flashes presented at the nine perceived speaker locations and lasted for the same 200-msec. duration as the auditory stimuli. Stereo disparity alone is routinely used to convey depth information accurately (Blakemore, 1970; Schor & Tyler, 1981; Bülthoff, Fahle, & Wegmann, 1991; Kontsevich & Tyler, 1994). The yellow (CIE $x = 0.43$, $y = 0.54$, 200 cd/m^2 calibrated luminance) DP104 phosphor CRT is specifically designed for stereoscopic applications, decaying to 0.1% peak intensity within 0.6 msec. The CRT was combined with fast-switching, large viewing-aperture, FE-1 ferroelectric shutter goggles to produce almost zero cross-talk in alternate frame stereoscopic presentations. In synchrony with the monitor, the goggles were driven by Cambridge Research Sys-

tems Visual Stimulus Generator VSG 2/5. This system allowed the presentation of stereoscopic flashes of light whenever the sound was presented and also provided a small (i.e., 0.1°) square image that could be moved in 3D space to locate the sounds. Ensuring millisecond timing coordination between the visual and auditory outputs required calibration using an Agilent 54641A two-channel Oscilloscope 350 MHz, 2 GSa/s.

Design

Observers were positioned on a chin rest throughout the experiment. Before the first condition, observers were blindfolded and sounds from each of the nine locations were recorded within the observers' inner ears. The nine speakers were placed in rows of three (one midline, and one left/right by 13.5°), separated by 20 cm in depth (Fig. 1; open square symbols).

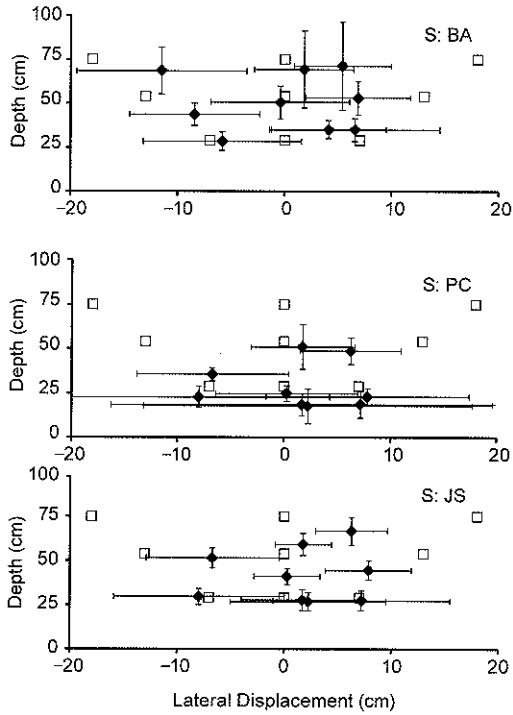

FIG. 1. Open black squares represent actual speaker locations as seen from above; coordinates 0,0 reflect the observer's nose, while 80 cm in depth reflects the CRT surface. Nine speaker locations, where 0,0 is the nose of the observer and 80 cm in depth is the depth plane of the CRT screen. Position 1 = far left; position 2 = far center; Position 3 = far right; Position 4 = middle left; Position 5 = middle center; Position 6 = middle right; Position 7 = near left; Position 8 = near center; Position 9 = near right. Diamonds are mean blind judgments for three observers to the sound-only condition (error bars are one *SD*).

The experiment was conducted in the same room as the sound recordings and contained a sound-only and an experimental condition. In the sound-only condition, observers randomly heard the nine sound recordings over headphones and used a joystick to move the stereovision square marker to the perceived sound location. Observers completed 81 trials during each of five 20-min. sessions, producing 45 judgments for each sound location. The experimental condition randomly presented a light and sound simultaneously and required observers to locate the sound using the square marker controlled by the joystick. Each observer had a unique set of auditory and visual stimuli locations. Observers completed 243 trials in each of 16 1-hr. sessions, yielding 48 responses for each auditory and visual location pairing. This condition used the same sound recordings as the sound-only condition and flashed the light at the location that the observer perceived the sounds to be located. It is important to note that the visual stimuli were *not* presented where the speakers were originally located in the free-field, but where observers perceived the sound to be located during the sound-only condition. This methodology makes the two perceived conditions coherent as the stimuli's physical locations could not be discerned accurately because the visual stimuli occurred at two (stereo) locations on the surface of the CRT. As such, the perceived locations of the auditory stimuli were deemed more appropriate than external physical measurements.

Of note, it would have been impossible to use LEDs and speakers in a 3D experiment as there would be no way to point to a location in depth without obstruction from physical equipment (i.e., the real speakers and LEDs in the front would get in the way of the LEDs and speakers in the back). Thus, use of the stereodisparity array and joystick to move a spot of light to the (virtual) location in space without obstructions between the observer and the CRT screen was chosen. An alternative option for making depth measurements, such as magnitude estimations, would present limitations greater than our virtual methodology.

Results

Fig. 1 shows the three observers' average placement of the stereo square-marker for the perceived locations of the nine speakers in the blind condition. Observers were somewhat accurate in their responses, although some sounds were perceived at locations notably different from the original speaker placements. Observer BA perceived the virtual locations in a pattern similar to the free-field locations, but with judgments on the left being more accurate than those on the right. Observer PC had reasonable localization in azimuth but perceived many of the virtual sounds closer to his face than expected. Finally, Observer JS also produced a pattern mostly representative of the speakers' locations, but with a few points, such as

those farthest from the observer in depth, moved away from their expected positions.

When considering the difficulty in localizing nearby sounds without a visual reference (Zahorik, 2001) and the close proximity of speakers' locations, this data set confirms the effectiveness of the virtual acoustic technology. However, it is important to acknowledge there is still a large amount of variability in these blind judgments and patterns of accuracy were not consistent across observers, with Observer BA demonstrating greater variability in depth judgments, Observer PC having more variability in azimuth judgments, and Observer JS having roughly equal variability in each dimension (see Fig. 1, where error bars report the variance in each dimension).

What is novel in this experiment is the use of virtual acoustics to perceptually match sounds apparently emanating from perceived locations in space with lights also apparently originating from given places in space. One would hope that this approach would reduce variability of judgment. There was some evidence of decreased variability in the experimental condition compared to the blind condition. Data in Table 1 show that Observer BA significantly decreased variability for azimuth (X) and depth (Z) judgments when auditory stimuli were accompanied by visual stimuli in the same location, Observer PC had decreased azimuth variability and increased depth variability, and Observer JS had similar variability in both conditions.

TABLE 1

MEAN VARIANCES AND STANDARD DEVIATIONS FOR THREE OBSERVERS BY CONDITIONS (DEG.)

Observer	Sound Only X		Coincident X		Sound Only Z		Coincident Z	
	M	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
BA	3.26	0.78	2.21	1.13	7.78	2.08	5.47	1.77
PC	2.22	1.23	1.77	0.62	5.39	3.85	6.94	2.21
JS	4.26	1.98	4.27	1.45	12.18	4.69	11.83	2.43

Fig. 2 expresses a sample of each observer's spatially coincident light and sound conditions from the experimental condition (only three of the nine possible positions are shown for brevity, i.e., Positions 1 (far depth, far left), 5 (middle depth, center), and 9 (near depth, far right). The data shown are representative of all observers' responses, with the three locations chosen because they are representative of how observers responded to each of the three depth planes. Of note, a presentation of all responses would require 90 graphs and is not feasible. Open squares indicate the speakers' actual location, while the intersection of the error bars (standard deviation) from the blind condition indicates both where the virtual sound was perceived and thus where the visual stimulus was presented, and diamonds indicate an observer's auditory localization. The title of

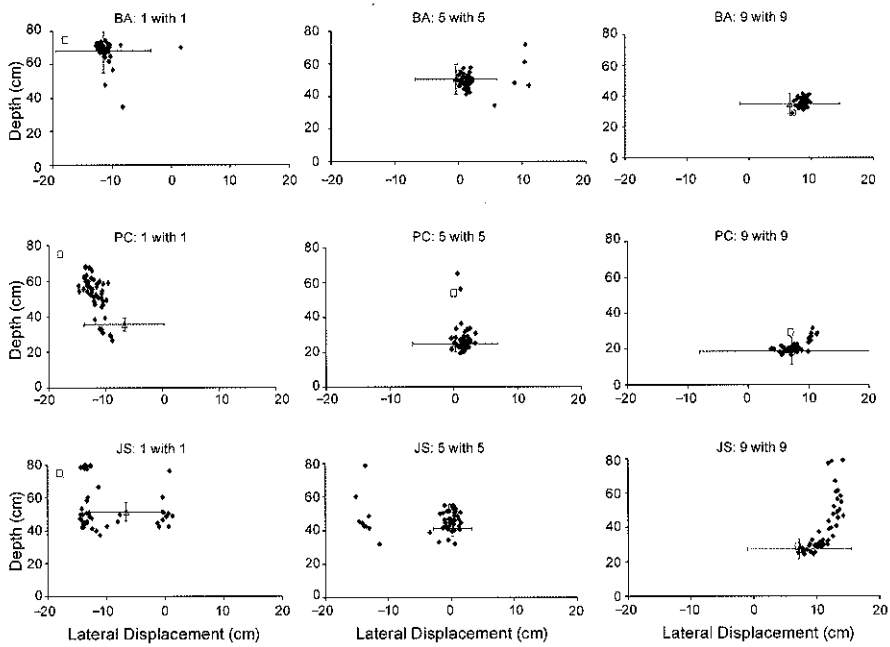

FIG. 2. Representative spatially coincident conditions for three observers

each graph details which sound location is played with which visual location. For instance, "1 with 1" means the first sound location with the first visual position, etc.

When the visual and auditory stimuli are presented from the same perceived location, observers were expected to and generally did exhibit consistency in their responses. For instance, Observer BA was highly consistent and accurate in judging each sound's location in the coincident condition, with the vast majority of responses located in the immediate proximity of stimuli presentation (*n.b.*, the intersection of the error bars from the blind condition indicates the location of the auditory signal derived from the previous blind condition, which is where the light was flashed). High accurate and consistent response was demonstrated by Observer PC for Positions 5 and 9, although this observer overshot the perceived stimuli location for Position 1. Different qualitatively, the data for Observer JS shows three distinct patterns. He produced three clusters of data for the first location, with one on each side of the stimulus location and one group near the computer monitor in depth but closer to the original speaker location in azimuth. Position 5 yielded a large group of responses appropriate to the positions of the stimuli, with only a few judgments displaced to the left. Responses for Position 9 indicated a different pattern altogether,

with judgments spread from the location of stimuli presentation to the far right of the field. In all, presenting a sound and light in the same position resulted in notably consistent and accurate auditory localization for two observers and a higher variability for the third.

To examine the effects of visual stimuli on auditory signals, data for each observer when the nine light locations (i.e., intersection of the error bars) were presented concurrently, with only the first auditory position, (i.e., open square) are used as a representative data set (Figs. 3, 4, and 5). Quantification of all the data follows these individual observers' samples. The first sound position paired with the first light position is the only spatially coincident case, with each other graph reflecting a noncoincident occurrence. As previously noted, Observer BA reliably located the sound at the appropriate position when the light and sound were presented together in space (Fig. 3, 1 with 1). When a light was presented at Positions 2 and 3 in the back row, Observer BA continued to perceive most of the sounds as emanating from their original location, independent of the light (Fig. 3). Although the light did draw a few judgments closer (i.e., toward the inter-

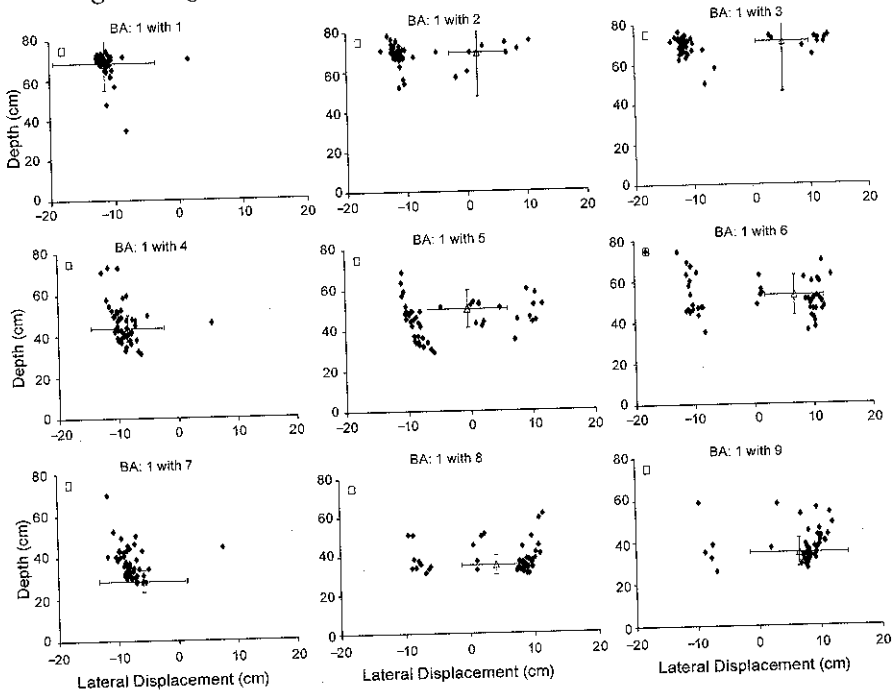

FIG. 3. Representative spatially noncoincident condition for S: BA. Open black squares indicate actual speaker's location, while open triangles and standard deviation error bars from the blind condition indicate light flash location, and diamonds indicate the observer's auditory localizations.

section of the error bars) and away from the sound source, there is virtually no consistent visual bias in the data. Note that the localizations fall close to the *virtual* auditory presentation (i.e., the intersection of the error bars present in" "1 with 1") and not near the actual speakers' location from the blind condition (i.e., the open square). For light Positions 4, 5, and 6, Observer BA brought the majority of judgments closer in depth to the middle row. Light Position 4 biased most of the judgments away from the virtual sound (i.e., falling very near the location of the light), while Light 5 led to a shift in localization which moved the majority of positions closer to the observer and to a similar depth as the light. There appear to be two groupings of responses for the fifth condition, with localizations falling to either side of the light. Presenting the stimuli at Position 6 also resulted in a noticeably bimodal distribution of responses, with many of the responses influenced only in depth by the light. Position 7 caused a high and consistent visual bias that clustered most of the points around the presentation of the light. Lights 8 and 9 were located close together and produced similar results, with nearly all responses biased to the depth of the visual stimuli and most clustered in azimuth around the light. There is a slightly stronger bimodal tendency in the data for Position 8, which may be expected given its closer azimuth location to speaker Position 1. Overall, Observer BA demonstrated some level of visual bias in all conditions, with bias being stronger for the light locations farthest away from the sound.

As seen in Fig. 4, Observer PC has a different pattern of responses than Observer BA. The former consistently overshot the presented light in the spatially coincident condition (*n.b.*, location of the intersection of the error bars in "1 with 1"), but otherwise shows strong biasing effects. The presence of Light 2 resulted in a bimodal distribution with a smaller group of responses located toward the left of the light and a larger group clustered around the light. This pattern is not found with Position 3, as this location produced a scatter of responses to either side of the light. Lights 4 and 6 show almost mirror-image patterns of bias, with responses located in a spread around the visual stimulus. Strong visual bias is seen at Position 5, although there is a separate distribution of points that are located at the light's depth and leftward of the azimuth location. Positions 7, 8, and 9, located closest to the observer, demonstrate almost 100% bias. In each of these conditions, the responses are clustered in small groups near the site of visual stimuli presentation, although Position 7 is shifted notably leftward. The responses of Observer PC are different in nature from those of Observer BA, but similarly exhibit a high amount of bias by each visual stimulus.

As with the observer's spatially coincident data, responses for Observer JS proved to be the most variable (Fig. 5). Light Positions 1, 2, 4, 5,

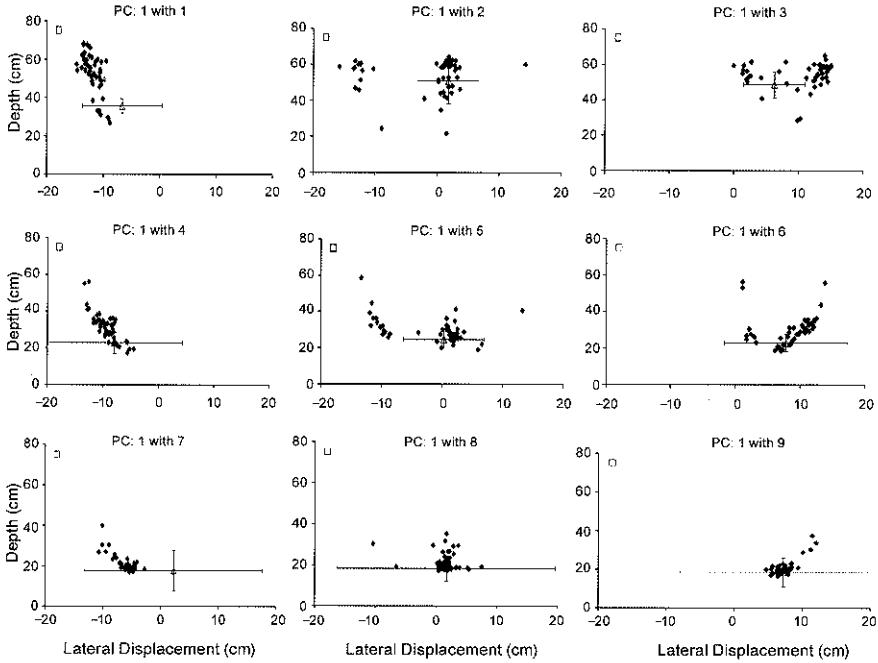

FIG. 4. Representative spatially noncoincident condition for S: PC. Open black squares indicate actual speaker's location, while open triangles and standard deviation error bars from the blind condition indicate light flash location, and diamonds indicate the observer's auditory localizations.

6, and 8 each resulted in a bimodal distribution of points at similar locations in front of and to the left of the observer, with these localizations positioned to the left of the virtual sound as measured from the blind condition (n.b., the location of the intersection of the error bars in "1 with 1"). The light in Position 3 seemed to have no discernable effect on auditory localization, as the distribution of data points did not yield a pattern suggesting a different interpretation than when the light was in Position 1 (i.e., compare "1 with 1" to "1 with 3"). Light 7 brought some responses closer to the visual stimuli but replicated elements of the grouping patterns found with other positions. The ninth light position produced the most bias, with a large number of points brought toward the light, but it is important to note that a significant number of responses were not biased and remained near the original perception of the auditory stimulus. The results of Observer JS do not contain the easily recognizable patterns of visual bias seen in Observers BA and PC.

A prominent characteristic of the ventriloquism effect is its variability across observers. For example, in two dimensions, Hairston, Vaughan,

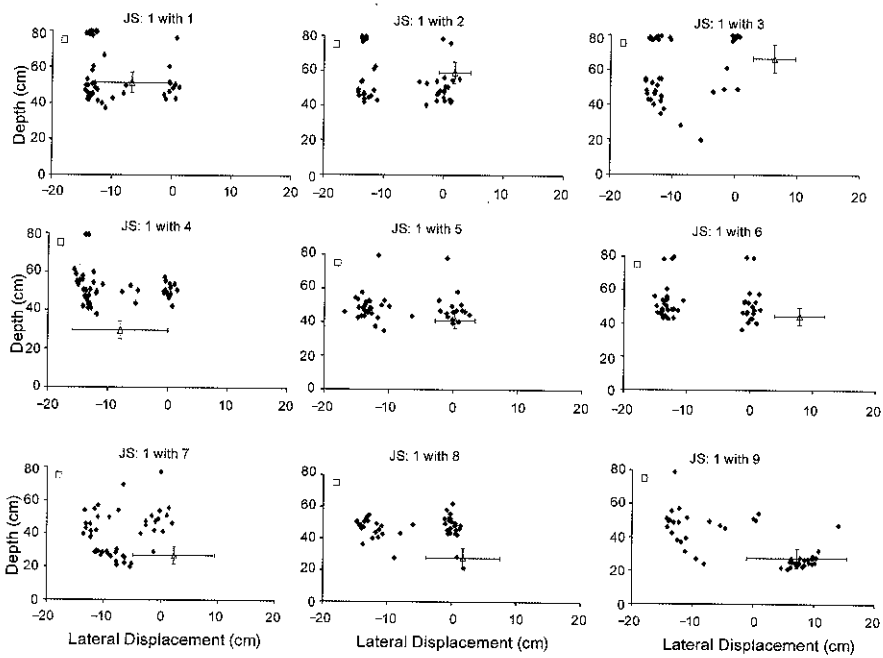

FIG. 5. Representative spatially noncoincident condition for S: JS. Open black squares indicate actual speaker's location, while open triangles and standard deviation error bars from the blind condition indicate light flash location, and diamonds indicate the observer's auditory localizations.

Wallace, Stein, and Schirillo (2001) found one observer who showed a consistently accurate localization of the sound source and produced nearly 0% visual bias, while another observer had almost 100% bias as he repeatedly perceived the sound as originating from the location of the light. The above data sets reflect this large inter-subject variability. Such variance in localization complicates the study of multisensory integration and must be taken into account when analyzing observers' performance, but is common throughout the literature. However, it is important to note that observers exhibit a relatively high consistency across their own judgments (Hairston, 2001).

To estimate the relation of visual bias with variability in locating sounds, the data for each observer were quantified across all responses and defined the term "bias" using the formula: $\text{Bias} = [(S_c - S_n)/(L_c - L_n)] \times 100$. The numerator provides a measure of the extent that localization is biased from the sound as S_c indicates the average location of the observer's response to the multisensory signal in the coincident case (light and sound at the same position) and S_n indicates the location of the observer's

response to the multisensory signal during the noncoincident case. In the denominator, L_c indicates the location of the sound at the coincident location and L_n indicates the location of the sound at the noncoincident location. Of note, the sound locations for L_c and L_n also represent the location of the presented lights as defined by this methodology. In essence, the flash of a noncoincident light occurred at the perceived location of the sound, which actually corresponded to the location of a noncoincident physical light. As previously stated, this was done to have both sounds and lights occur in perceptual, not physical, coordinates.

Fig. 6 depicts the visual bias obtained in each dimension (i.e., azimuth = X, depth = Z) for each observer in each of the nine sound locations. For example, Observer BA's percent bias in the depth (Z) direction (i.e., hatched +45° bars) was greater than his percent bias in the azimuth (X) direction (i.e., hatched –45° bars) when the sound was played in Position 1 (i.e., 89% bias and 64% bias, respectively). Each of these two data points represents the average amount of visual bias incurred from all eight (noncoincident) positions from which the light was flashed and demonstrates that bias was greater in depth than in azimuth. Visual bias was also significantly greater in depth when averaged across all nine sound locations for all three observers (i.e., designated as the mean black and gray at the right of the graph, indicating 57% bias in the depth direction but only 35% bias in the azimuth direction). Analysis of variance (ANOVA) shows that this pattern was statistically significant for all three observers (Observer BA: $F_{1,8} = 60.97$, $p < .0001$, Cohen's $d = 3.57$; Observer JS: $F_{1,8} = 10.86$, $p < .01$, Cohen's $d = 3.28$; Observer PC: $F_{1,8} = 5.23$, $p < .05$, Cohen's $d = 0.93$). However, it is clear that percent bias differs across observers and conditions and reversals of this trend are present (e.g., see Observer BA's sound Location 8). There are also cases of negative bias, when the sound was located in the direction *opposite* that of the noncoincident light (e.g., see Observer JS's sound Location 8 and the fact that for Observers JS and PC the y-axis scale depicts negative bias).

Also plotted on Fig. 6 is the amount of variance (in degrees of visual angle) determined in the sound-only condition obtained in each dimension (i.e., azimuth = X, depth = Z) for each observer in each of the nine sound locations (n.b., values on the second y-axis). These are shown as hatched bars, with the –45° hatched bars reflecting variance in the azimuth (X) dimension and the +45° hatched bars reflecting variance in the depth (Z) dimension. The observers' data showed different relations, as Observer BA showed more variance in the depth dimension than in the azimuth dimension (i.e., 11° and 6°, respectively). Observer JS continued this but was less, and Observer PC had more variance in the azimuth than depth

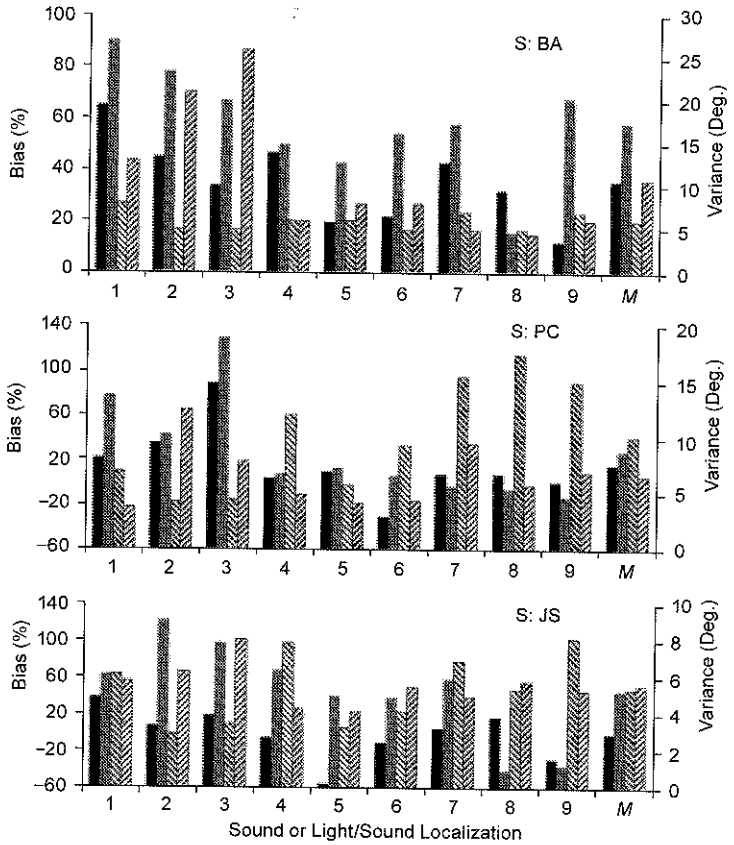

FIG. 6. Percent bias obtained in azimuth (X, black bars) and depth (Z, gray bars) dimensions for each observer for each of the nine sound locations. Variance (in degrees visual angle) obtained in azimuth (X, hatched −45° bars) and depth (Z, hatched +45° bars) dimensions for each observer for each of the nine and mean sound locations (values on second Y-axis).

dimension. To explore the relationship between visual bias and variance, a mean difference score was computed between bias in the Z and X directions and a difference score between variance in the Z and X directions and the correlation between these differences was plotted for each observer (Fig. 7).

Table 2 depicts performance times from the onset of the stimuli until the observer completed auditory localization. It shows a positive correlation for all three observers (BA: $r^2 = .12$; JS: $r^2 = .16$; PC: $r^2 = .18$), indicating that a relative increase in the variability of depth judgments in the sound-only condition was related to a similar increase in bias. In this manner, it appears that variability and bias were stronger in depth than in azimuth

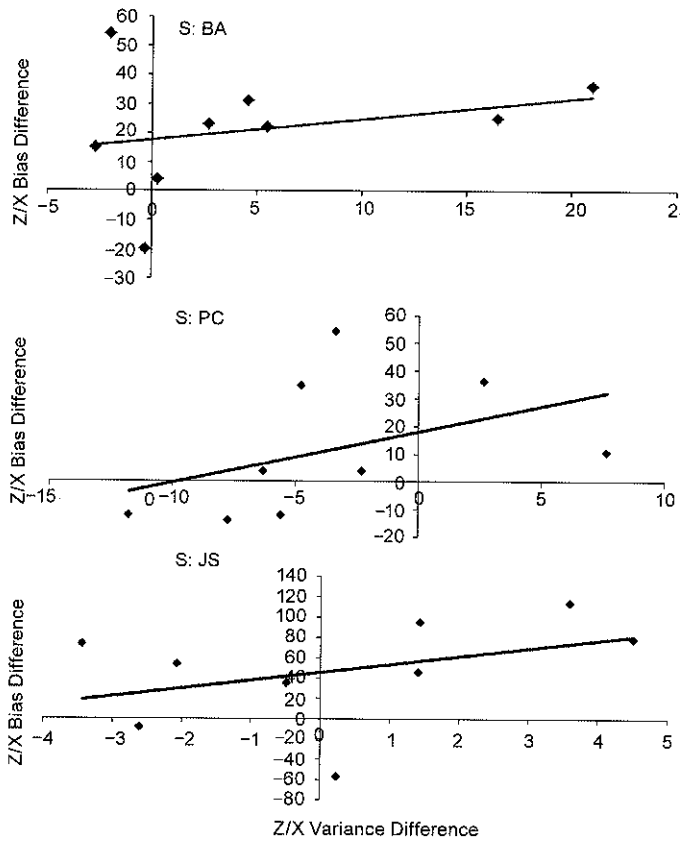

FIG. 7. Correlation of bias and variance difference scores for each of nine sound locations.

across all observers. The fact that this occurs only in the majority of data points reflects that the Bayesian model is probabilistic.

Observers' performance times (computed from the onset of the sound/light until the observer pulled the joystick trigger to judge the final localization of the sound) were affected by condition. Table 2 depicts the performance time in msec. for the sound only condition, the spatially noncoincident auditory-visual condition, and the spatially coincident auditory-visual condition. For all three observers, two-tailed t-ratios indicated that both multisensory conditions reflected statistically significant reductions ($p < .001$) in performance times compared with the sound-only condition (for the coincident cases, $df = 431$: Observer BA: $t = 11.97$, Cohen's $d = 2.05$, 95%$CI = \pm 14.32$; Observer PC: $t = 8.60$, Cohen's $d = 4.84$, 95%$CI = \pm 7.89$; Observer JS: $t = 15.02$, Cohen's $d = 2.21$, 95%$CI = \pm 24.84$; for

TABLE 2
MEAN PERFORMANCE TIMES FOR THREE OBSERVERS BY CONDITIONS (MSEC.)

Observer	Coincident		Noncoincident		Sound Only	
	M	*SD*	*M*	*SD*	*M*	*SD*
BA	2,157	49	2,445	172	3,262	32
PC	4,111	27	3,995	66	5,678	37
JS	4,096	85	4,422	95	7,747	217

Note.—Performance time from stimulus onset until observer completed auditory localization.

the noncoincident cases, $df = 431$: Observer BA: $t = 9.21$, Cohen's $d = 6.60$, 95%$CI = \pm 50.25$; Observer PC: $t = 8.96$, Cohen's $d = 3.15$, 95%$CI = \pm 19.28$; Observer JS: $t = 12.78$, Cohen's $d = 1.99$, 95%$CI = \pm 50.25$). In other words, performance times were dramatically shorter when visual and auditory information were present than when auditory stimuli were presented alone. This is important as the increase in performance time (e.g., between 2.0 and 7.5 sec.) could increase the variability in localizations. Moreover, the noncoincident performance time was statistically significantly greater than the coincident performance time for Observers BA and JS ($df = 431$; Observer BA: $t = 5.39$, $p < .001$, Cohen's $d = 2.28$, 95%$CI = \pm 9.35$; Observer JS: $t = 1.97$, $p < .05$, Cohen's $d = 3.62$, 95%$CI = \pm 10.81$). Observer PC's data did not support this as differences between the conditions fell short of statistical significance (Observer PC: $t = 1.12$, $p = .26$, Cohen's $d = 2.30$, 95%$CI = \pm 63.4$).

DISCUSSION

Given the difficulty of auditory localization without a visual cue, the large variance obtained in these observers was expected (Zahorik, 2002b). Yet, despite this variability, the data from the blind condition indicate a moderate success in producing three-dimensional externalized sounds (Fig. 1), and the virtual auditory simulation seems a reasonable means of stimulus presentation (Moller, 1992; McKinley, Erickson, & D'Angelo, 1994; Langendijk & Bronkhorst, 1999; Schilling & Shinn-Cunningham, 2002).

One noteworthy difference in the blind localization judgments compared to previous work is the absence of overestimations in the near field. Although Zahorik (2002a) found observers consistently overestimated distances within 3 m, present observers (Fig. 1) consistently *underestimated* the location of virtual acoustic stimuli. There are several possible explanations for this discrepancy. In his experimental design, Zahorik blindfolded observers prior to testing to prevent any prior knowledge of the testing environment. Although there also were observers blindfolded while making recordings, observers did have previous knowledge about the size and general layout of the testing environment. Additionally, these ob-

servers made judgments with open eyes, meaning there was more information available to make relative distance judgments (Ashmead, *et al.*, 1990). Based on available environmental information in the present experiment, it may have been inappropriate for observers to overestimate the farthest speaker locations as such judgments would be based beyond the visual boundary of the CRT screen. A combination of these factors likely led observers to underestimate rather than overestimate near-field auditory depth.

Overall, the experimental condition produced many instances of visual bias (Fig. 6). Present results support the prior findings of Gardner (1968) and Mershon, *et al.*, (1980) that localization in depth is strongly biased by visual stimuli while being incongruent with Zahorik's finding (2002c) that visual stimuli enhance auditory localization under similar conditions. Also, a good amount of intersubject variability was noted and expected as prior two-dimensional multisensory research has shown bias can vary between 0 and 100% and reflect different qualitative patterns among observers (Hairston, 2001; Hairston, Vaughan, Wallace, Stein, Norris, & Schirillo, 2003).

Some of the data seem incompatible with earlier 2-D research which indicates visual bias was more likely when stimuli were located *closer* together in space (Hairston, *et al.*, 2003; Zampini, Guest, Shore, & Spence, 2005). For example, the row of lights with the most consistent bias in Fig. 3 is located nearest to Observer BA and further away in depth from the sound source (i.e., pairing the sound with Lights 7, 8, or 9), while there is hardly any bias for Light 2 or 3 presented in the same depth plane as the sound. However, as visual bias has not previously been studied in depth, results from earlier research may apply only to bias in azimuth. Thurlow and Jack (1973) showed more "capture" in the vertical plane when stimuli were presented in a reverberant room compared to when stimuli were presented in a sound-absorbent room and claimed that reverberations and other spectral cues should *decrease* variability of localization. In contrast, the hypothesized and observed *increase* in variability of localization produced *more* visual bias. This suggests that the characteristics of visual bias may differ in depth, which reflect in sound localization uses of different cues to make azimuth and depth judgments. Although conclusions are significantly limited by this small sample size, preliminary support was found for the hypothesis that greater variance in depth compared to in azimuth would result in additional visual bias (Table 1). Bayesian analysis contends that this theoretical framework is valid, although the finding is the first known which concerns this selectively in the depth dimension (Battaglia, *et al.*, 2003; Alais & Burr, 2004; Rowland, *et al.*, 2007).

In agreement with the aforementioned studies, making judgments in the spatially coincident compared to the noncoincident conditions re-

duced localization performance time, measured from the stimuli presentation to observers' localization with the joystick, for two observers (Table 2). Moreover, for all three observers any multimodal signal (whether coincident or not) resulted in a dramatic reduction in performance time. Not to be confused with reaction times, performance times are quite long as observers have to manipulate the joystick for several seconds to localize the brief auditory stimuli, and speed of response was never mentioned to the observer.

One concern with the current experimental paradigm is that the visual stimuli occurred at the perceived, instead of the actual, location of the auditory stimuli. Ideally, the actual and perceived auditory locations should be similar, but this was not always the case. For example, Observer PC had a few cases wherein there were relatively large spatial differences between the actual speakers' locations and blind responses (Figs. 2 and 4). Although this makes interpreting bias problematic, all multimodal signals used the perceived location of the unimodal sound as the light source, and results seem to provide support for using the perceived auditory and visual locations as congruent.

It is well accepted that concurrently experiencing a spatially noncoincident light and sound increases the probability that the sound will be located closer to the light. Though stimulus characteristics affect the extent of multisensory bias, these patterns are also highly variable across observers (Hairston, *et al.*, 2001). While the ventriloquism effect has strong empirical support in one dimension, it has seldom been studied in depth and has only now been addressed simultaneously across two dimensions.

To examine the ventriloquism effect in two-dimensional space, it was necessary to create an experimental paradigm capable of presenting stimuli in front of observers without obstructing their view or providing visual references. To accomplish this goal, virtual acoustic and stereo-imaging techniques were combined to study multisensory bias in an important and novel way not previously attempted in the literature.

The experimental results show similarities to one-dimensional multisensory bias, and also indicate novel and significant differences. Observers' responses varied, but exhibited consistent patterns within individual datasets. For a given observer, patterns of visual bias found in two-dimensions were often similar to the bias previously found in one-dimension. In general, lights may bias the location of a sound completely, bias the sound partially to its location, or have no effect. However, present data suggest that there is a general increase in variance for sound location in depth compared to in azimuth. Moreover, even with large interobserver differences, overall mean visual bias was similarly greater in depth (Fig. 6). While this relationship is modest, it is important to realize that it is con-

sistent across all observers despite their very different patterns of bias and variance. Although it is difficult to make generalizations from the current data, this is the initial report of a difference in the amount of visual bias obtained in the depth dimension versus azimuth. Its correlation with the variance produced in locating sounds suggests that a Bayesian interpretation of visual bias holds across both dimensions.

REFERENCES

Alais, D., & Burr, D. (2004) The ventriloquist effect results from near-optimal bimodal integration. *Current Biology,* 14, 257-262.

Ashmead, D. H., LeRoy, D., & Odom, R. (1990) Perception of the relative distances of nearby sound sources. *Perception & Psychophysics,* 47, 326-331.

Battaglia, P. W., Jacobs, R. A., & Aslin, R. N. (2003) Bayesian integration of visual and auditory signals for spatial localization. *Journal of the Optical Society of America A,* 20, 1391-1397.

Bertelson, P., & de Gelder, B. (2004) The psychology of multimodal perception. In C. Spence & J. Driver (Eds.), *Crossmodal space and crossmodal attention.* Oxford, UK: Oxford Univer. Press. Pp. 151-177.

Bertelson, P., & Radeau, M. (1981) Cross-modal bias and perceptual fusion with auditory-visual spatial discordance. *Perception & Psychophysics,* 29, 578-584.

Blakemore, C. (1970) The range and scope of binocular depth discrimination in man. *Journal of Physiology,* 211, 599-622.

Bronkhorst, A. W., & Houtgast, T. (1999) Auditory distance perception in rooms. *Nature,* 397, 517-520.

Bülthoff, H., Fahle, M., & Wegmann, M. (1991) Perceived depth scales with disparity gradient. *Perception,* 20, 145-153.

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

Durlach, N. I., Shinn-Cunningham, B. G., & Held, R. M. (1993) Supernormal auditory localization. *Presence,* 2, 89-103.

Hairston, W. D. (2001) The influence of vision on auditory localization. Unpublished master's thesis, Wake Forest Univer.

Hairston, W. D., Vaughan, J. W., Wallace, M. T., Stein, B. E., Norris, J. L., & Schirillo, J. A. (2003) Visual localization ability influences cross-modal bias. *The Journal of Cognitive Neuroscience,* 15, 20-29.

Hairston, W. D., Vaughan, J. W., Wallace, M. T., Stein, B. E., & Schirillo, J. A. (2001) Cross-modal bias is related to variability in visual localization. Presented at the Society for Neuroscience, 13 November, San Diego, CA.

Gardner, M. B. (1968) Proximity image effect in sound localization. *Journal of the Acoustical Society of America,* 43, 163.

Guski, R. (1990) Auditory localization: effects of reflecting surfaces. *Perception,* 19, 819-830.

Kontsevich, L. L., & Tyler, C. W. (1994) Analysis of stereo thresholds for stimuli below 2.15 c/deg. *Vision Research,* 34, 2317-2329.

Langendijk, E. H. A., & Bronkhorst, A. W. (1999) Fidelity of three-dimensional-sound reproduction using a virtual auditory display. *Journal of the Acoustical Society of America,* 107, 528-537.

LINDAU, A., MAEMPEL, H. J., & WEINZIERL, S. (2008) Minimum BRIR grid resolution for dynamic binaural synthesis. *Journal of the Acoustical Society of America,* 123, 3498.

MCKINLEY, R. L., ERICKSON, M. A., & D'ANGELO, W. R. (1994) 3-dimensional auditory displays: development, applications, and performance. *Aviation, Space, and Environmental Medicine,* 65, A31-A38.

MERSHON, D. H., BALLENGER, W. L., LITTLE, A. D., MCMURTRY, P. L., & BUCHANAN, J. L. (1989) Effects of room reflectance and background noise on perceived auditory distance. *Perception,* 18, 403-416.

MERSHON, D. H., DESAULNIERS, D. H., AMERSON, T. L., & KIEFER, S. A. (1980) Visual capture in auditory distance perception: proximity image effect reconsidered. *Journal of Auditory Research,* 20, 129-136.

MOLLER, H. (1992) Fundamentals of binaural technology. *Applied Acoustics,* 36, 171-218.

RIFE, D. D., & VANDERKOOY, J. (1989) Transfer-function measurement with maximum-length sequences. *Journal of the Audio Engineering Society,* 37, 419-444.

ROWLAND, B. A., STANFORD, T., & STEIN, B. (2007) Multisensory integration shortens physiological response latencies. *Journal of Neuroscience,* 27, 5879-5884.

SAVIOJA, L., HUOPANIEMI, J., LOKKI, T., & VÄÄNÄNEN, R. (1990) Creating interactive virtual acoustic environments. *Journal of the Audio Engineering Society,* 4, 675-705.

SCHILLING, R. D., & SHINN-CUNNINGHAM, B. (2002) Virtual auditory displays. In K. M. Stanney (Ed.), *Handbook of virtual environments: design, implementation, and applications.* Mahwah, NJ: Erlbaum. Pp. 65-92.

SCHOR, C. M., & TYLER, C.W. (1981) Spatiotemporal properties of Panum's fusional area. *Vision Research,* 21, 683-692.

THURLOW, W. R., & JACK, C. E. (1973) Certain determinants of the ventriloquism effect. *Perceptual and Motor Skills,* 36, 1171-1184.

VESA, S., & LOKKI, T. (2006) Detection of room reflections. *Proceedings of the 9th International Conference on Digital Audio Effects (DAFx-06),* Montreal, Canada, September 18–20.

WAGNER, H. (2004) A comparison of neural computations underlying stereo vision and sound localization. *Journal of Physiology–Paris,* 98, 135-145.

WALLACE, M. T., ROBERSON, G. E., HAIRSTON, W. B., STEIN, B. E., VAUGHAN, J. W., & SCHIRILLO, J. A. (2004) Unifying multisensory signals across time and space. *Experimental Brain Research,* 158, 252-258.

WARREN, D. H., WELCH, R. B., & MCCARTHY, T. J. (1981) The role of visual-auditory "compellingness" in the ventriloquism effect: implications for transitivity among the spatial senses. *Perception & Psychophysics,* 30, 557-564.

WENZEL, E. M., ARRUDA, M., KISTLER, D. J., & WIGHTMAN, F. L. (1993) Localization using nonindividualized head-related transfer functions. *Journal of the Acoustical Society of America,* 94, 111-123.

WIGHTMAN, F. L., & KISTLER, D. J. (1989a) Headphone simulation of free-field listening: I. Stimulus synthesis. *Journal of the Acoustical Society of America,* 85, 858-867.

WIGHTMAN, F. L., & KISTLER, D. J. (1989b) Headphone simulation of free-field listening: II. Psychophysical validation. *Journal of the Acoustical Society of America,* 85, 868-878.

ZAHORIK, P. (2001) Estimating sound source distance with and without vision. *Optometry and Vision Science,* 78, 270-275.

ZAHORIK, P. (2002a) Assessing auditory distance perception using virtual acoustics. *Journal of the Acoustical Society of America,* 111, 1832-1846.

ZAHORIK, P. (2002b) Auditory display of sound source distance. Presented at the Proceedings of the 2002 International Conference on Auditory Display, Kyoto, Japan.

ZAHORIK, P. (2002c) Direct-to-reverberant energy ratio sensitivity. *Journal of the Acoustical Society of America,* 112, 2110-2117.

ZAMPINI, M., GUEST, S., SHORE, D. I., & SPENCE, C. (2005) Audio-visual simultaneity judgments. *Perception & Psychophysics,* 67, 531-544.

Accepted November 12, 2010.