

APRIL 03 2002

Assessing auditory distance perception using virtual acoustics

Pavel Zahorik



J Acoust Soc Am 111, 1832–1846 (2002)
<https://doi.org/10.1121/1.1458027>



CrossMark

Related Content

Auditory bubbles reveal sparse time-frequency cues subserving identification of musical voices and instruments

J Acoust Soc Am (October 2016)

Two issues at the interface between speech and music

J Acoust Soc Am (November 2006)

Auditory and Visual Sensations: Yoichi Ando's theory of architectural acoustics.

J Acoust Soc Am (October 2010)

Downloaded from http://pubs.aip.org/asa/jasa/article-pdf/111/4/1832/808939/1832_1_online.pdf



**Advance your science and career
as a member of the**
ACOUSTICAL SOCIETY OF AMERICA

LEARN MORE



Assessing auditory distance perception using virtual acoustics

Pavel Zahorik^{a)}

Waisman Center, University of Wisconsin-Madison, Madison, Wisconsin 53705

(Received 27 March 2001; revised 14 January 2002; accepted 14 January 2002)

In most naturally occurring situations, multiple acoustic properties of the sound reaching a listener's ears change as sound source distance changes. Because many of these acoustic properties, or cues, can be confounded with variation in the acoustic properties of the source and the environment, the perceptual processes subserving distance localization likely combine and weight multiple cues in order to produce stable estimates of sound source distance. Here, this cue-weighting process is examined psychophysically, using a method of virtual acoustics that allows precise measurement and control of the acoustic cues thought to be salient for distance perception in a representative large-room environment. Though listeners' judgments of sound source distance are found to consistently and exponentially underestimate true distance, the perceptual weight assigned to two primary distance cues (intensity and direct-to-reverberant energy ratio) varies substantially as a function of both sound source type (noise and speech) and angular position (0° and 90° relative to the median plane). These results suggest that the cue-weighting process is flexible, and able to adapt to individual distance cues that vary as a result of source properties and environmental conditions.

© 2002 Acoustical Society of America. [DOI: 10.1121/1.1458027]

PACS numbers: 43.66.Qp, 43.66.Pn, 43.66.Yw [LRB]

I. INTRODUCTION

Compared to directional localization, relatively little is known about human perception of sound source distance. It is clear that for a complete description of perceived three-dimensional auditory space, distance is no less important than direction. Rigorous study of auditory distance perception is often difficult, however, because of inherent complexities of the stimulus. Changes in sound source distance often result in changes to multiple acoustic properties of the sound reaching a listener's ears, properties which themselves may be affected by factors other than source distance. In many cases, properties of the sound source itself and of the acoustic environment may be confounded with the acoustic changes resulting from distance changes. The perceptual processes subserving distance localization therefore likely analyze and sensibly combine multiple sources of acoustic information, or cues, in order to produce stable estimates of sound source distance. Perhaps a result of the potentially unreliable nature of distance cues, the accuracy with which sound source distance is judged has often times been found to be quite poor (Békésy, 1949; Bronkhorst and Houtgast, 1999; Coleman, 1968; Holt and Thurlow, 1969; Loomis *et al.*, 1998; Mershon *et al.*, 1989; Nielsen, 1993).

At least four possible acoustic distance cues have been proposed for conditions where both listener and sound source are stationary (Mershon and King, 1975). These may be briefly described as follows.

Intensity. In general, sound intensity at the listener's position decreases when the distance between listener and fixed-power sound source is increased. The precise nature of this intensity change, however, depends on both environmental characteristics and various properties of the sound source, including acoustic power and radiation patterns. Under ideal

conditions (a point source in an acoustic free field), intensity loss as a function of distance obeys an inverse-square law, which implies a 6-dB intensity loss for each doubling of distance. Notable departures from this function occur when the acoustic environment contains surfaces that reflect sound. In these conditions, the rate of loss of intensity as a function of distance is decreased. It is also important to note that source distance and source power are completely confounded in measures of intensity at the ear. As a result, the auditory system is presumably forced to rely on assumptions about source power in order to use intensity reliably as a distance cue. This fact has led, in part, to the suggestion that the accuracy of distance judgments may be improved if the sound source is familiar to the listener (Coleman, 1962; McGregor, Horn, and Todd, 1985), because this familiarity may aid in assumptions relating to the acoustic power of the source.

Direct-to-reverberant energy ratio. In environments with sound reflecting surfaces, the ratio of energy reaching a listener directly (without contact with reflecting surfaces) to energy reaching the listener after reflecting surface contact (reverberant energy) decreases systematically with increases in source distance. In rooms, change in direct-to-reverberant energy ratio results primarily from the effect of the inverse-square law on the direct portion of the sound field, because the energy in the later-arriving reflected portion of the sound field may be well approximated by a diffuse sound field, which is defined to have uniform energy over varying source positions. For a given room, reverberant energy as a function of time is determined principally by the size of the room and the acoustic properties of the reflecting surfaces of the room. Many outdoor environments also produce reverberation, and hence a direct-to-reverberant energy ratio cue that varies with distance (Richards and Wiley, 1980).

Spectrum. At least two circumstances cause systematic

^{a)}Electronic mail: zahorik@waisman.wisc.edu

changes in the at-the-ear spectrum as a function of distance. For distances greater than approximately 15 m (Blauert, 1983), the sound-absorbing properties of air significantly modify the sound source spectrum. In general, these absorbing properties of air attenuate high frequencies the most, although the effect is relatively small: on the order of 3- to 4-dB loss per 100 meters at 4 kHz (Ingard, 1953). A second type of spectral change occurs in sound-reflective environments where the spectrum that reaches the ear may be affected by the acoustic properties of the reflective surfaces. As distance increases, the proportion of reflected energy increases, thereby potentially changing the at-the-ear spectrum systematically. Like the intensity cue, spectral cue changes with distance are confounded with changes in the sound source spectrum. As a result, sound source familiarity may also enhance the utility of this cue.

Binaural differences. When sound sources are in the acoustic near-field, binaural differences in both intensity and time are no longer independent of radial distance, as they are for far-field planar waves. These differences, often referred to as differences resulting from acoustic parallax, are maximal along the interaural axis, and decrease to zero on the median plane. For example, Hartley and Fry (1921) have shown that interaural intensity differences (IID) for a sound source (1860-Hz sinusoid) on the interaural axis can differ for distances between 87.5 and 17.5 cm by as much as 20 dB (values derived using a spherical head model). Interaural time differences (ITD) changes with distance have been shown to be less salient than those for IID (Brungart and Rabinowitz, 1999; Duda and Martens, 1998).

Among the most important unanswered questions regarding the perception of sound source distance concerns the way in which information from these multiple acoustic cues is sensibly combined and processed. Because of the complexity of the stimulus, past experiments have typically chosen to manipulate only one of these acoustic cues at a time, while either removing or holding constant all other cues. Although this approach retains strict experimental control, the resulting stimulus may bear little resemblance to that encountered in natural situations, in which multiple distance cues are available to the listener. The recent advances in virtual acoustic technology (Møller *et al.*, 1995; Wightman and Kistler, 1989) make it possible to avoid this trade-off, as precise stimulus control is available under stimulus conditions that are ostensibly identical to those occurring in natural environments. As such, the use of virtual acoustic technology is ideally suited for examining listeners' use of multiple distance cues in a natural environment.

This article describes a series of experiments that examines listeners' abilities to judge sound source distance in a room environment, and specifically the ways in which listeners combine and weight multiple distance cues. Virtual acoustic technology is extensively used, both as a method for precisely quantifying the acoustic stimulus reaching the listeners' ears, and as a means of realistic and accurate stimulus presentation. The rest of the article is divided into three sections. The first (Sec. II) describes the measurement procedure that quantifies the at-the-ear stimulus for a variety of distances in a representative room environment. These measure-

ments are then evaluated in terms of potential acoustic cues to distance. Section III describes an experiment in which psychophysical distance functions are measured for virtual sound source stimuli constructed from the individualized measurements described in Sec. II. These stimuli allow all of the acoustic distance cues present in the real room environment to vary as they would naturally. The last section (Sec. IV) describes an experiment in which certain distance cues present in the virtual sound source stimulus are manipulated simultaneously. This experiment addresses the relative salience of various acoustic distance cues in a room environment.

II. ACOUSTICAL ANALYSIS OF POTENTIAL SOURCE DISTANCE CUES

One of the major limitations of past work on auditory distance perception has been the quantification and control of the stimulus reaching the listener's ears. To overcome this limitation, the current work employs modern virtual acoustic techniques to measure impulse responses of an acoustic system that contains: a listener, a room environment thought (*a priori*) to be rich in acoustic distance cues, and a sound source that can be varied in distance from the listener. The methods used to measure these impulse responses, which will be referred to as *binaural room impulse responses* (BRIRs), are fundamentally similar to those used for anechoic measures (e.g., Møller *et al.*, 1995; Wightman and Kistler, 1989), but require longer excitation periods to fully capture the response of the room environment. Because these impulse responses are essentially complete representations of the proximal stimulus reaching the listener's two ears, all acoustic distance cues are represented within them and may therefore be precisely quantified.

A. Methods

1. Environment

A small auditorium (264-person seating capacity) located within the Waisman Center at the University of Wisconsin—Madison served as the test environment. This room was chosen based on the following characteristics: (a) its size; (b) its reverberation characteristics; and (c) its availability for the purposes of this experiment. The auditorium had a total volume of approximately 830 m³ and total surface area of approximately 653 m². Its shape was complex, with sloping floor (−6°) and ceiling (11°), as well as nonparallel sections of the side walls. The main floor (not including stage area) was approximately rectangular: 14 m long by 12.2 m wide. The majority of floor surfaces were carpeted, excluding seating areas and the stage, which were covered in tile and wood, respectively. The ceiling was covered with acoustical tile and the walls were composed primarily of painted drywall material.

The sound source consisted of a small, high-quality loudspeaker (Realistic Minimus 3.5) with a 90-mm full-range driver. This loudspeaker was mounted on a tripod that could be easily moved up and down one aisle of the auditorium. Twelve distances extending from a fixed origin point near the end of one aisle were examined: 0.30, 0.43, 0.61, 0.86, 1.22, 1.72, 2.44, 3.45, 4.88, 6.90, 9.75, and 13.79 m.

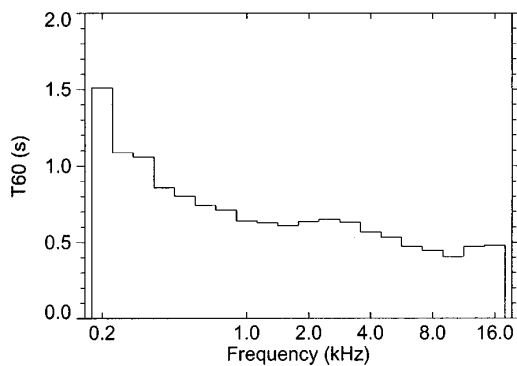


FIG. 1. Reverberation time, T_{60} , for the test environment (a medium-sized auditorium) as a function of frequency. Results have been averaged across the 12 distances measured with a reference microphone.

The loudspeaker was always oriented with its driver facing the origin. In order to accommodate the farthest distance, the loudspeaker had to be placed on the auditorium's stage (approximately 1.5 m from the front edge). At this position, the loudspeaker height was decreased in order to account for the 0.38-m height of the stage. For all other positions, no attempt was made to adjust the loudspeaker height to compensate for the floor slope.

Figure 1 displays measured reverberation time (T_{60}) as a function of frequency for this environment. These measurements were made with a reference microphone (Brüel & Kjær $\frac{1}{2}$ -in. microphone, cartridge type 4133) using methods similar to those described by Schroeder (1965).

2. Participants

Nine paid volunteers (8 female and 1 male, ages 20–28) participated in the binaural impulse-response measurements. None of the participants had any prior exposure to the auditorium where the measurements took place. These participants also served as listeners in all subsequent experiments.

3. Measurement technique and apparatus

Binaural room impulse responses (BRIRs) were measured using methods fundamentally similar to those described in detail by Møller *et al.* (1995), although much longer excitation periods were used to capture the response of the room. Miniature electret microphones (Sennheiser KE4-211-2) were inserted into each participant's ear canals such that the microphone diaphragm was at the position of the canal entrance, and an acoustic seal was formed between the microphone and canal wall using Etymotic Research ER-13R-2 ring seals (a blocked-meatus configuration). Two impulse-response measurements were made with the loudspeaker positioned (ear height for the 0.30-m source, approximately 1.25 m from the floor surface) at each of the 12 distances. For one measurement the seated participant faced the loudspeaker (0° azimuth), for the other the listener was rotated 90° such that the loudspeaker was opposite the right ear (90° azimuth). All measurements were made with the loudspeaker driven with a fixed power level that produced approximately 80 dB SPL at 0.30 m with the participant removed.

A technique using a maximum-length sequence (MLS) excitation signal, a type of pseudorandom noise, was used to measure all impulse responses (Rife and Vanderkooy, 1989). Generally, this technique excites the system under evaluation periodically with an MLS signal, while recording the results and then averaging these results periodically. The system's impulse response is computed by cross correlating the averaged response with the raw MLS. This technique offers a number of advantages over past impulse-response measurement techniques in terms of excitation signal generation, stability of required computations, and signal-to-noise ratio (Rife and Vanderkooy, 1989). In situations where nonstationary noise disturbances exist, a weighted average of responses to the excitation signal periods may be used to further improve signal-to-noise ratio. Nielsen (1998) has shown that weighting response periods by the reciprocal of the mean-squared level in each period can substantially improve signal-to-noise ratio in conditions of high-level nonstationary disturbances. This weighted averaging procedure was used for all impulse responses in this application.

Specifically developed software was used to implement the MLS technique on Tucker–Davis Technologies (TDT) hardware (DD1 2-channel D/A and A/D, AP2 array processor card with optical link to the DD1) with two simultaneous recording channels and the ability to average recording periods in real time. For all measurements, a 15th-order MLS (32 767 points) was used. D/A and A/D conversion was effected with 16-bit precision at a sampling frequency of 50.0 kHz. Forty-four periods of the MLS signal were presented, with the results of the final 40 periods averaged (weighted) in real time. The responses to the first four periods were not recorded in order to insure steady-state system excitation for the final 40 periods. Microphone outputs were fed first to custom-built amplifiers, then into TDT MA2 microphone amplifiers to provide an additional 10 dB of gain prior to TDT DD1 A/D input. The output of the DD1 D/A was fed to a Crown D-75 amplifier with fixed gain driving the movable loudspeaker. The computed impulse responses were stored in floating-point format on a Pentium-class PC used to control the TDT hardware.

It should be noted that a different technique for impulse-response measurement was initially utilized in the present study. This method, using a Golay code stimulus (Foster, 1986; Zhou and Green, 1992) was found to have serious shortcomings when used to measure systems that are to a small degree time-variant, such as binaural recordings from humans. These disturbances, caused by such factors as small head movements, resulted in impulse-response artifacts when using the Golay technique (Zahorik, 2000). Given these problems, the Golay technique was abandoned in favor of the MLS technique.

Headphone (Beyerdynamic DT 990 Pro) impulse responses when coupled to the ears of a given participant were also measured. These measurements were collected for the generation of virtual sound sources to be used in subsequent experiments. Since it is important that the microphone position in the ear canal is the same for both BRIR measurements and these ear-coupled headphone IR measurements, both were obtained during the same measurement session. Param-

eters of the MLS recording system were different in this application, because the headphone IRs are much shorter than the room IRs. Two hundred ten repetitions of a 9th-order MLS (511 points) were presented at a 50-kHz sampling rate, averaging the final 200 repetitions. The playing/recording equipment chain was the same as for the BRIR measurements, except for a TDT headphone buffer (HB1) and 40 dB of attenuation provided by a Hewlett-Packard 350C attenuator placed between the D/A and headphone buffer. In order to assess headphone placement variability on the participant's ears, three replication measurements were collected in which the participant removed and replaced the headphones between replications.

Reference measurements were also made for each of the 12 distances with the participant absent, using a Brüel & Kjær $\frac{1}{2}$ -in. condenser microphone (omnidirectional cartridge type 4133) placed at the position normally occupied by the center of the participant's head, oriented to point to the loudspeaker source. These measurements were used to both evaluate system noise, as well as for level calibration purposes. The microphone's amplified output (Brüel & Kjær model 2609 amplifier) was fed to the TDT DD1 A/D. Impulse-response measurements were made for each distance using the MLS system previously described. Noise measurements were also collected by simply turning off the excitation signal to the loudspeaker. As a result, the noise measurements reflect the average of 40 recording periods, having, in theory, a mean level of $10 \log(40)$ dB less than an unaveraged noise measurement.

4. Procedure

Each participant was met outside the auditorium and briefed on the measurement procedure. After placing microphones in the ears, the participant was blindfolded and led into the auditorium, where he/she was seated in a chair positioned at the zero point. Participants were instructed to remain as quiet and motionless as possible during the course of a given measurement. If the operator detected any noticeable participant motion or unusual environmental noise during a measurement, that measurement was discarded and another collected at the same position. Measurements were obtained from the 12 distances and 2 angular orientations followed by the series of headphone measurements. Upon completion of measurements, participants were led out of the auditorium and blindfolds removed. Hence, the participants were at no time allowed to examine the auditorium visually. Auditory experience with the acoustic environment was also minimized, given the attenuation provided by the sealed-canal microphone placement. The total time required to complete the measurement for each participant was approximately 30 min.

5. Post-processing

Signal-to-noise ratio present in the room was determined from the reference measurements. Figure 2 displays the magnitude spectra of measured impulse responses for three instances of room excitation: measurement signal at the closest distance (0.30 m), measurement signal at the farthest dis-

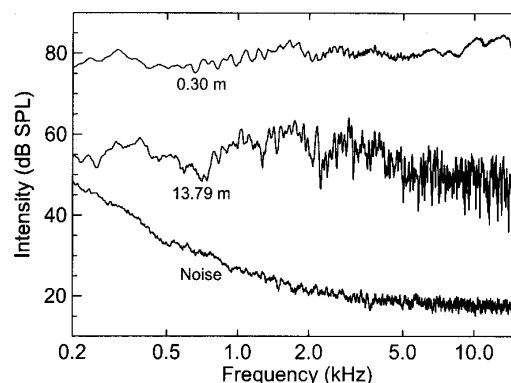


FIG. 2. Magnitude spectra of the sound source signal at 13.79 m, signal at 0.3 m, and the noise alone as measured with a reference microphone and averaged 40 times.

tance (13.79 m), and no measurement signal (noise alone). The y-axis scale is based on an observed average sound pressure of 80 dB SPL at the closest distance. The noise is greatest at the low frequencies, which is caused primarily by the air-handling system in the auditorium. As a result, signal-to-noise at the low frequencies was poorest, with the smallest ratio—on the order of 7 dB at 200 Hz—existing when the signal is farthest away. Over much of the frequency range, however, signal-to-noise ratios even for the farthest signal are greater than 30 dB. The closest signal produces the best signal-to-noise ratios: 45 to 60 dB over the range of 0.5 to 10 kHz. It should be emphasized that these are *averaged* results. The actual SPL of the unaveraged noise was approximately 16 dB higher.

Poor signal-to-noise ratio was, therefore, a substantial issue in the present measurements, especially at low frequencies for distant sound sources. Two standard approaches for increasing signal-to-noise ratio were not practical in this situation. Increases in excitation level would have risked nonlinear loudspeaker operation as well as potential participant fatigue/discomfort. Significant increases in the number of excitation signal-averaging periods would have caused increases in measurement time, and as a result increased the chances of measurement contamination resulting from participant motion. An alternative method of post-processing the BRIR measurements was therefore used in an attempt to minimize noise contamination in regions of poor signal-to-noise ratio, particularly the later portions of the response where much of the reverberant energy has decayed away. This noise contamination may be seen as a change in slope of the cumulative energy decay functions (Schroeder, 1979). An example of this is shown in Fig. 3, where the slope of the measured decay (dark curve) becomes substantially more shallow at around 200 ms.

Because both the noise and the diffuse reverberant signal have essentially Gaussian probability-density functions (independent in each ear), the noise may be treated as though it were simply reverberation with a different (and incorrect) decay rate. Imposing the proper decay rate onto the noise will in theory result in a close approximation to an uncontaminated impulse response. This decay rate correction must

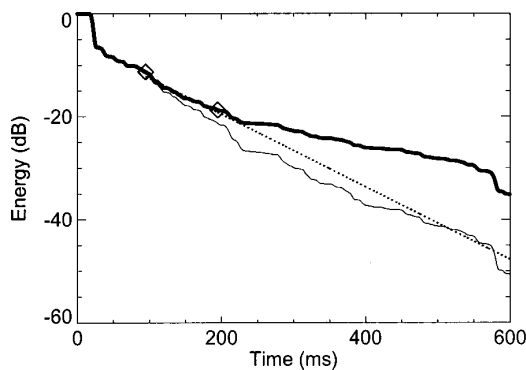


FIG. 3. Energy decay function (dark curve) for the frequency band centered at 635 Hz (participant SRQ, left ear, 4.88-m source distance). The dotted line is a linear fit to the decay function region between the two diamond symbols. The light curve is the derived decay function after application of the exponential decay time-windowing procedure.

be done in a frequency-dependent manner, however, because reverberation decay rate is known to be frequency dependent (as seen in the T60 measure of Fig. 1).

To implement this style of noise reduction, BRIRs were first passed through a bank of bandpass filters (zero-phase FIR) with 1/3-octave-band center-frequency spacing from 0.2 to 16 kHz. Cumulative energy decay curves were then computed for each frequency band, followed by a linear fit of the decay curve in dB, from 80 to 180 ms after the direct-path arrival. The dotted line in Fig. 3 shows the linear fit in dB to the decay curve in the region between the diamond symbols (80 to 180 ms post direct-path arrival). An exponentially decaying time window was then constructed based on the slope of this linear fit and applied to the corresponding impulse response from each frequency band. The light curve in Fig. 3 shows the decay curve for the windowed result. For 1/3-octave bands ≤ 400 Hz, slopes of the fitted decay functions were increased by 30% in order to ensure that most low-frequency components of the impulse response would decay sufficiently within its 655-ms period. Windowed results from all frequency bands were then summed in the time domain to recompose a *windowed* version of the original impulse response. Figure 4 displays decay contours as a function of frequency before [Fig. 4(a)] and after [Fig. 4(b)] application of this windowing procedure. It may be seen that the windowing procedure is effective at removing the later nonlinearly decaying (in dB) noise contamination. All subsequent use of BRIRs refer to measurements that have been modified by this post-processing stage.

6. Potential limitations

At least two potential limitations of the described methods for measuring and processing the BRIRs should be noted. First, the 1/3-octave filtering and reconstruction operations used in this post-processing procedure did impose a degree of ripple to the frequency spectrum of the processed result, due to the overlapping passbands in the filters. Second, the measurement and subsequent post-processing both caused distortion of the low-frequency decay of reverberant energy. The measurement period should ideally have been

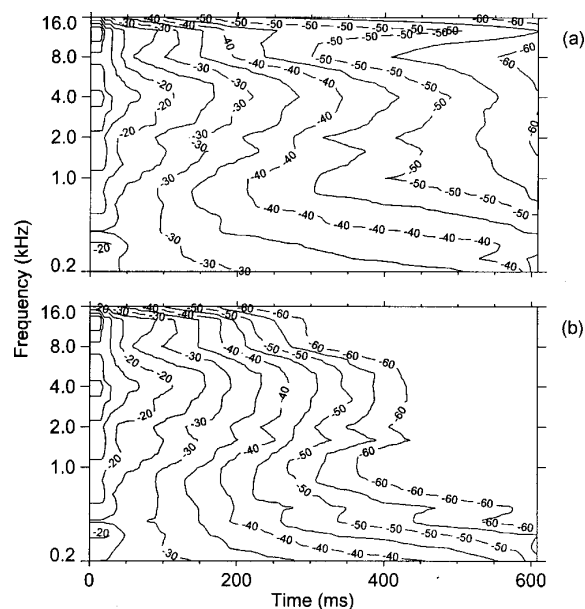


FIG. 4. Energy decay function contours for a BRIR measured from 4.88 m (participant SRQ, left ear) before (a) and after (b) application of the frequency-dependent time-windowing procedure.

extended to fully capture all low-frequency decay components, although technical and noise-level limitations prevented this. In an attempt to minimize these effects, the post-processing stage was specifically designed to increase the rates of reverberant energy decay below 400 Hz.

Both of these limitations were not believed to significantly impact the current application, principally because direct psychophysical comparisons were not made between the real and simulated environments. The amount of frequency ripple imposed by the filterbank was both reasonably small (within ± 3 dB) and constant across all processed BRIRs. It could therefore be easily attributed to a property of the sound source itself, such as a loudspeaker response. Similarly, the artificially increased decay of low-frequency reverberation can be thought to represent an environment that was slightly more damped in the low frequencies than the physical measurement environment. Clearly, other applications may require further attention to these issues.

B. Results

Analyses of potential acoustic cues to distance were performed on the post-processed BRIRs, for each of the 12 distances and 9 participants. For each analysis, results were averaged across participants ($n=9$), in order to focus on acoustical aspects that change with source distance change, and to remove the small acoustical idiosyncrasies resulting from individual participant characteristics.

1. Intensity: Monaural and binaural

Figure 5 displays relative intensity changes as a function of source distance for the two angular orientations (0° and 90° azimuth). Intensity was computed based on each entire

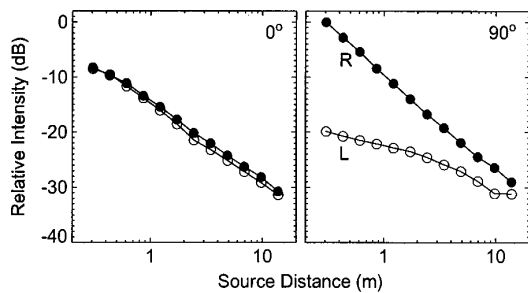


FIG. 5. Average ($n=9$) relative intensity change as a function of source distance for angular orientations of 0° (left panel) and 90° (right panel) azimuth. Solid symbol curves show data for the right (R) ear. Open symbols show data for the left (L) ear. Intensity decay rate for L and R ears at 0° is approximately -4 dB per doubling of source distance. Decay rate for L and R ears at 90° is approximately -2 and -5 dB per doubling distance.

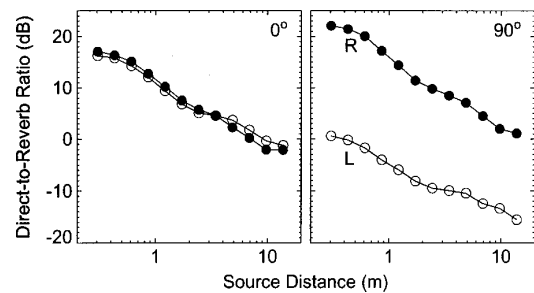


FIG. 6. Average ($n=9$) direct-to-reverberant energy ratio change as a function of source distance for angular orientations of 0° (left panel) and 90° (right panel) azimuth. Solid symbol curves show data for the right (R) ear. Open symbols show data for the left (L) ear. Direct-to-reverberant energy ratio decay rate for all cases is approximately -3 to -4 dB per doubling source distance.

BRIR. As a result, the relative intensity values represent both direct-path energy as well as reverberant energy, explaining why the intensity change per distance doubling in all cases is less than the 6 dB predicted by an inverse square law. The binaural differences in intensity change between the left and right ears for the 90° orientation may also be explained by reverberant energy. As the source distance is increased, direct-path energy, which causes the large intensity differences between the two ears for the 90° orientation, will decrease. Therefore, the diffuse reverberant energy, which excites both ears equally, will account for a greater proportion of the total energy at the farther distances. For the closest distances, additional near-field effects may also contribute to the interaural intensity differences at 90° .

From these results, it is likely that relative intensity is a salient distance cue for this environment. Intensity changes were both systematic and quite large—changes for this environment that corresponded to a doubling of source distance were on the order of 3 to 5 times best estimates of an intensity jnd (just-noticeable difference) (Jesteadt, Wier, and Green, 1977). Binaural differences in intensity for source directions displaced from the median plane can also provide listeners with additional distance information.

2. Direct-to-reverberant energy ratio

The ratio of direct to reverberant energy was computed for each of the BRIRs by a simple time-windowing procedure. The start of the direct-path portion of the impulse response was first determined by finding the first time at which energy was greater than 20 dB above the noise floor. Because anechoic (direct-path only) head-related transfer functions are known to have impulse-response lengths on the order of 2.5 ms (Møller *et al.*, 1995), the direct-path portion of the BRIRs was defined to be 2.5 ms in duration. The reverberant portion of the BRIRs was defined to be the remainder of the IR following the 2.5-ms direct-path portion. Direct-to-reverberant energy ratio is simply the ratio of total energies in each of these regions of the BRIRs.

This time-window duration was effective at separating the direct path from reflections for all but the farthest sound sources where the delay between the direct path and the floor

reflection was very small. Given the absorption properties of the carpeted floor, the energy contribution of the floor reflection was quite small, however.

Figure 6 displays direct-to-reverberant energy ratios as a function of source distance for the two angular orientations (0° and 90° azimuth). The decrease in ratios is due primarily to the decrease in direct-path energy (which decreases by 6 dB per distance doubling), given that reverberant energy is acoustically diffuse, and therefore relatively constant within rooms. In the present room environment, reverberant energy was found to decrease by approximately 1 dB per doubling of source distance.

Although the observed changes in direct-to-reverberant energy appear to be substantial, there are at present no relevant psychophysical results (i.e., no jnd measures) with which to compare this variation.

3. Spectrum

Change in spectral shape of the BRIRs with varying distance was examined via a 1/3-octave-band analysis, which was the same as that implemented for measurement post-processing. Figure 7 displays average magnitude spectra ($n=9$) for the two ears and two angular orientations. In order to remove effects of overall level (see Fig. 5), each spectrum

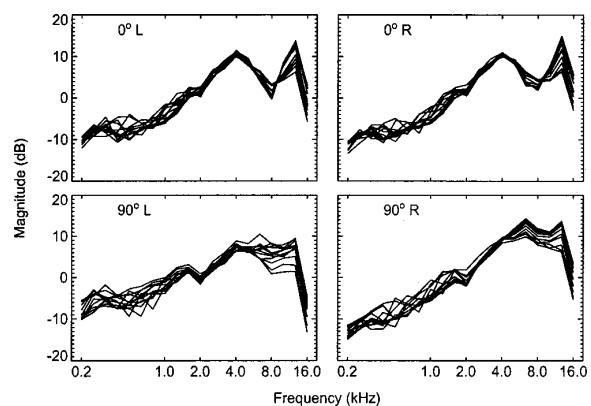


FIG. 7. Average ($n=9$) 1/3-octave-band BRIR spectral shape with the mean magnitude for each distance removed. Azimuth angle (0° or 90°) and ear (L or R) are displayed in each panel.

has been normalized by its mean magnitude. The only systematic changes in source spectrum with increasing distance are the slight decreases in high-frequency energy, which decrease by at most 9 dB over the full range of source distances, or roughly 1.6 dB per distance doubling. These small changes observed in spectral shape suggest that it is unlikely to be a salient distance cue in this situation, given known sensitivity to spectral shape changes under conditions where source distance is fixed (Green and Mason, 1985).

III. APPARENT SOURCE DISTANCE

This section describes an experiment in which psychophysical distance functions were estimated for nine listeners, using a direct scaling procedure on the virtual sound source stimuli constructed from the individualized measurements described in the previous section. All of the acoustic distance cues present in the real room environment were present in the virtual display, and varied as they would naturally. Because both source direction (Gardner, 1969; Holt and Thurlow, 1969; Mershon *et al.*, 1989), and familiarity (Coleman, 1962; McGregor *et al.*, 1985) with the source stimulus may provide listeners with additional distance cue information, these factors were manipulated in the experiment.

A. Methods

1. Participants

Nine paid volunteers (8 female and 1 male, ages 20–28) participated as listeners. All had participated previously in the binaural room impulse-response (BRIR) measurements, and had normal pure-tone thresholds, as verified by audiometric screening at 15 dB HL.

2. Stimulus generation and presentation

Virtual sound source stimuli were generated using the blocked-meatus BRIR and headphone impulse-response (IR) measurements collected for each listener (see Sec. II) following the technique described by Møller *et al.* (1995). As part of this technique, 32 768-coefficient FIR filters were constructed from the post-processed BRIR measurements for each individual listener ($n=9$) and spatial location: 12 distances (0.3–13.79 m) and 2 angular orientations (0° and 90°). Similarly, headphone equalization filters were constructed based on the individual headphone IRs. In an attempt to minimize the acoustic effects of headphone placement variability, an average (rms) of the three headphone measurement replications was computed based on each measurement's smoothed (20% of a critical bandwidth) magnitude spectrum. The averaged spectrum was then inverted and bandpass filtered between 0.2 and 16 kHz and defined to have linear phase, in order to construct a 256-coefficient finite impulse response (FIR) equalization filter. To simulate loudspeaker reproduction of a given input signal in an auditorium room environment for a given listener at a given spatial location, a BRIR filter for the given listener and location along with that listener's headphone equalization filters were applied to the input signal and presented over headphones (Beyerdynamic DT 990 Pro) using Tucker-Davis Technologies hardware (PD1) for D/A conversion (16-bit, 50 kHz).

Two types of source signals were used in these experiments: a noise burst and a speech syllable. The noise burst was Gaussian with 50-ms duration and was gated on/off with a cosine-squared window of duration 2.5 ms up/down. The speech sample was an anechoic recording of a male speaker uttering the syllable /da/ at a conversational level, 1 m from the recording microphone. Levels in the virtual display were adjusted such that the natural anechoic level of the speech syllable was preserved at a distance of 1 m.

3. Procedure

Listeners were seated in a sound-attenuating booth and asked to judge the apparent egocentric distance in feet of virtual sound sources presented in a random order via a direct scaling procedure. The unit of feet was chosen because all listeners, except listener SSI, indicated that they were most comfortable with reporting distances in British units. Listener SSI preferred SI units, and therefore responded in meters. A response of zero (feet or meters) was reserved for those sources that appeared to originate from a location within the listener's head. Listeners were instructed to close their eyes during the self-paced source presentation. Responses were made (with eyes open) on a computer terminal with a numeric keypad, using a minimum precision of two decimal places. Stimulus type (noise or speech) and source angular position (0° or 90°) were the same within a block of 120 trials (10 repetitions per source distance). Four blocks of trials, corresponding to the combinations of stimulus type and angular source position, were run in a single 2-h session.

B. Results

All listener's judgments of apparent source distance were transformed to meters and plotted as a function of virtual source distance for each of the four stimulus conditions: noise 0° , noise 90° , speech 0° , and speech 90° . These results, which are assumed to be good estimates of perceived distance, are displayed in Figs. 8–11, with separate panels in each figure for individual listeners. Within each panel, point symbols represent responses on individual trials. A small degree of random variation in the x dimension was introduced when displaying these data, so as to avoid symbol occlusion for like responses across trials. The square symbols are geometric means of responses for each distance. Linear fits (on logarithmic coordinates) to the geometric means are shown in each panel, along with their corresponding slopes in the lower right corner of each panel. No zero responses were reported, indicating that in all cases listeners perceived sources to be externalized from the head.

Since the proportion of variance accounted for by the linear fit on logarithmic coordinates is generally quite high (R^2 ranging from 0.7 to 0.9), it may be concluded that the data are well approximated by a compressive power function of the form

$$\psi_r = k \phi_r^a,$$

where ψ_r is perceived distance, ϕ_r is physical source distance, a is the power-law exponent, and k is a constant. On the logarithmic coordinates of Figs. 8–11, a is equivalent to

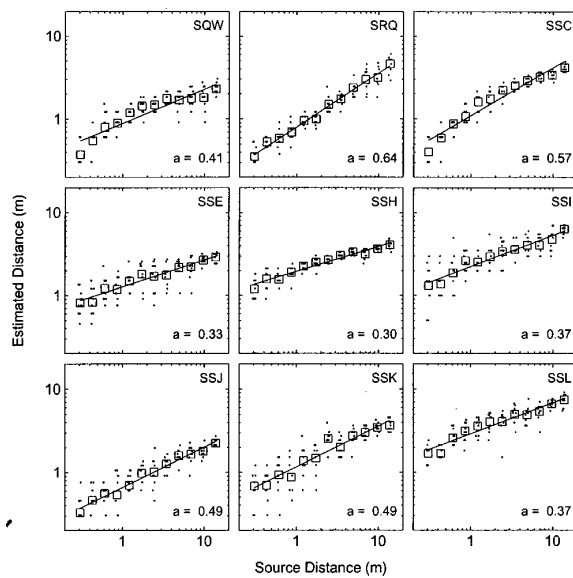


FIG. 8. Individual apparent distance judgments as a function of physical distance for the noise-burst stimulus at 0° azimuth. Each panel displays data for one listener. Point symbols represent individual trial data. The square symbols are geometric means. A linear fit to the data on log-log coordinates is also displayed, with slope (a) shown in the lower right.

the slope of linear fit, and $\log(k)$ the intercept. The average value of a was approximately 0.39 (s.d.=0.13) across all listeners and stimulus conditions, which is substantially less than the veridical value of 1.0. The average value of k was approximately 1.32 (s.d.=0.56), slightly higher than the veridical $k=1$. These results suggest that listeners tend to make large underestimates of source distance for most sources, and that the amount of underestimation is related exponentially to the actual source distance. Very close sources, however,

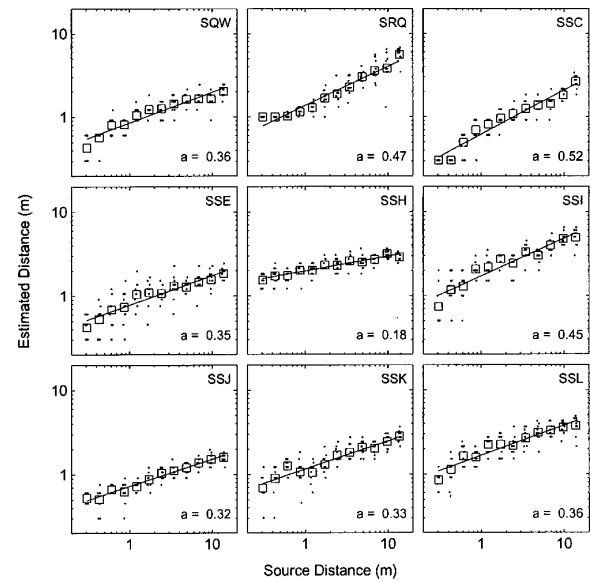


FIG. 10. Individual apparent distance judgments as a function of physical distance for the speech stimulus at 0° azimuth.

are actually (on average) overestimated, because values of k are greater than 1.

At least two lines of evidence suggest that these observed errors are in fact perceptual errors, and do not result merely from the method of response. First, similar response methods have been used to study visual distance perception that result in near-perfect accuracy under full-cue conditions (see Da Silva, 1985, for a summary of relevant literature). Second, for auditory distance, other response methods have been shown to produce results very similar to those collected using methods analogous to the direct scaling procedure implemented here (Loomis *et al.*, 1998).

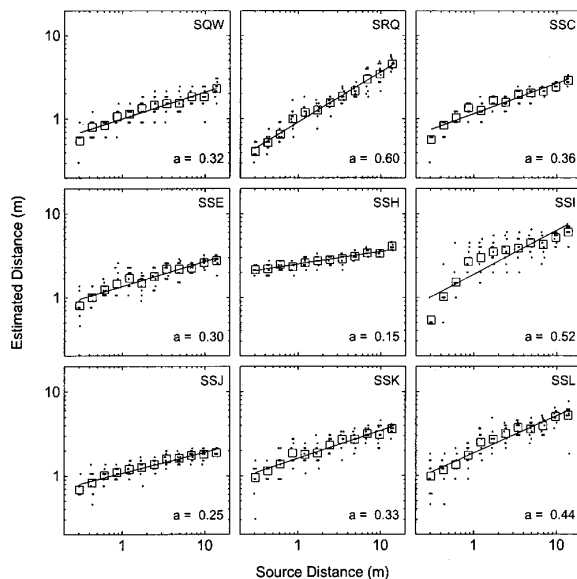


FIG. 9. Individual apparent distance judgments as a function of physical distance for the noise-burst stimulus at 90° azimuth.

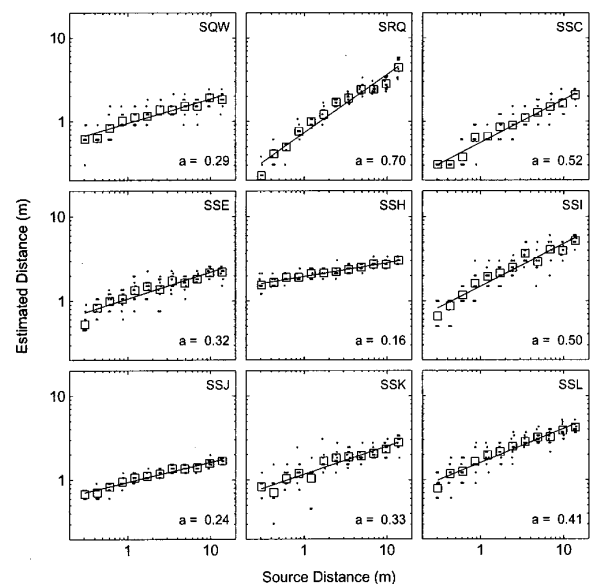


FIG. 11. Individual apparent distance judgments as a function of physical distance for the speech stimulus at 90° azimuth.

Although all listeners tended to underestimate most sound source distances, substantial differences existed in the amount of underestimation. The standard deviation of computed exponents within an experimental condition ranged from 0.10 to 0.17. This result is not surprising, given the known individual variability in judgment of visual target distance. Da Silva (1985) reports exponent standard deviations ranging from 0.09 to 0.18 for judgment of visual target distances using both magnitude estimation and ratio estimation methods.

Because consistent effects of angular orientation and stimulus type were not immediately observable in Figs. 8–11 as a result of individual listener variability, two repeated-measures analyses of variance (ANOVAs) were used to examine these effects more closely. One ANOVA had the fitted exponents, a , as its dependent measure. Since ANOVA has been shown to be robust with respect to violations of its normality assumption (Pearson, 1931), the fact that the a population is potentially non-normally distributed does not likely affect the outcome of the analysis. The other ANOVA had the constant values, $\log(k)$, as its dependent measure. Both ANOVAs treated angular orientation and stimulus type as within-subjects factors. In order to hold experiment-wise alpha level at the chosen criterion level of 0.05, alpha must be split for evaluation of each separate ANOVA, because the two dependent measures are not logically independent. No statistically significant differences between angular orientations or stimulus types were found for either dependent measure at the 0.025 significance level. The values of a were found to be significantly less than 1 on average, $F(1,8) = 257.89$, $p < 0.0001$; and the values of $\log(k)$ were found to be significantly greater than zero on average, $F(1,8) = 59.79$, $p < 0.0001$, indicating that k was significantly larger than the veridical value of 1.

An additional aspect of the data may be seen by visual inspection of Figs. 8–11. For many listeners, change in perceived distance appears to increase more rapidly for near sources (less than approximately 1.2 m) than for far sources (e.g., Fig. 9, listener SSI). The cause of these results is at present unknown, although it appears unlikely to be stimulus based. None of the physical cue measures reported in the previous section showed qualitatively different effects for near and far sources. Additionally, near-field binaural cues (which were not explicitly measured) were also unlikely to have caused these patterns of results because they were observed at source directions on the median plane for certain listeners (e.g., Fig. 8, listener SQW).

In summary, psychophysical distance functions were found to be well approximated by compressive power functions, suggesting that most source distances are underestimated. Although a substantial amount of individual variability was observed, the psychophysical functions do not appear to be systematically affected by either source signal type or source direction.

IV. LISTENER WEIGHTING OF ACOUSTIC CUES

The goal of this experiment is to evaluate the relative importance listeners place on individual acoustic distance cues under naturalistic stimulus conditions. Since multiple

acoustic parameters covary with physical sound source distance in most, if not all, naturally occurring environments, it is of interest to know how listeners combine the information from these multiple cues to arrive at a single distance percept. For the environment examined in this article, intensity and direct-to-reverberant energy ratio appear to be the acoustic parameters that vary most with source distance (see Sec. II). As such, listeners' use of these two cues is highlighted. Of additional interest is how cue use may change with varying source angle or source signal type.

This basic question, how it is that complex stimulus information is combined to form a single percept, has been the motivation for considerable study. Among the more famous early inquiries were those made by Brunswik (1952). This program of study involved correlating measures of naturally occurring physical parameters in the environment with perceptual responses, in an attempt to determine the perceptual weight placed on each physical parameter. Though innovative, this method experienced problems when all physical parameters were highly correlated with each other (causing problems of multicollinearity; Cohen and Cohen, 1983). More modern techniques experimentally apply small amounts of independent random perturbation to each physical parameter of interest (Berg, 1989, 1990; Berg and Green, 1990; Lutfi, 1995; Lutfi and Oh, 1997; Stellmack *et al.*, 1997; Willihnganz *et al.*, 1997). Because these perturbations are small, departures from natural stimulation are argued to be minimal. This is one of the principal advantages of this technique over more conventional factorial designs, which result in many stimulus presentations with highly unnatural combinations of cues.

In this technique, perceptual weights are estimated by examining the relation between responses and physical parameter perturbations. Estimation procedures using a maximum-likelihood criterion based on a presumed model (Ahumada and Lovell, 1971; Berg, 1989) and a least-squares criterion (Lutfi, 1995; Lutfi and Oh, 1997) have both been used. Because fewer assumptions are required by the latter procedure (Lutfi, 1995), it was used in the current experiments, correlating the trial-to-trial perturbation values for intensity and direct-to-reverberant energy ratio with apparent distance responses to determine the perceptual weight placed on each distance cue.

The experiment described here quantifies the relative perceptual weight listeners place on the intensity, I , and direct-to-reverberant energy ratio, ν , when making distance judgments under three different stimulus conditions. The first condition presented a noise-burst signal with equal degrees of perturbation applied to the naturally occurring values of I and ν cues at each distance. The second condition was the same as the first, except that the source signal was a speech syllable. The final condition attempts to verify that the magnitudes of perturbation applied to each cue do not affect the derived perceptual weights. In this condition, the noise-burst signal was presented with different degrees of perturbation were applied to I and ν . All stimulus conditions examined source positions at both 0° and 90° azimuth. Additionally, because it is possible that I and ν weights change with changing distance, these interactions were also evaluated.

A. Methods

1. Participants

Seven paid volunteers (all female, ages 20–28) participated as listeners in the first and second conditions of the experiment. Five of these listeners also participated in the third verification conditions. All listeners participated previously in the binaural room impulse-response (BRIR) measurements, as well as in the baseline experiments used to determine psychophysical distance functions.

2. Stimulus generation and presentation

Virtual sound source stimuli were constructed using methods previously described (see Sec. III), with individualized BRIRs (post-processed) and individualized headphone equalization filters used for all listeners. As before, BRIRs from 12 different distances and two angular orientations were used (see Sec. II). Perturbations of the I and ν cues were performed on each BRIR from trial to trial in the following way. Intensity was perturbed by scaling the BRIRs. Scale values were computed using samples drawn from a normal distribution with zero mean (relative to the given natural cue value at each distance) and $\sigma_I = 3$ dB. Direct-to-reverberant energy ratio, ν , was perturbed by scaling only the direct-path portion of the BRIRs, which was defined to extend from the first point in the impulse response with appreciable energy to a point 2.5 ms later (see Sec. II for details). These scale values were also computed using samples drawn from a normal distribution with zero mean. For the first two stimulus conditions, this distribution had $\sigma_\nu = 3$ dB. The final verification condition set $\sigma_\nu = 6$ dB. Perturbations values for I and ν were drawn independently, with each being applied equally to left- and right-ear BRIRs. It should be stressed that on any one trial, actual cue values for I and ν may be thought to consist of two parts: the nominal cue value present in each BRIR (see Figs. 5 and 6), and the additional perturbation value. Therefore, perturbation distribution means were effectively centered on the nominal values of the I and ν cues for each distance. The values for perturbation standard deviation, $\sigma_I = \sigma_\nu = 3$ dB, were chosen to span approximately a doubling of source distance (see Figs. 5 and 6). As such, the potential for wildly conflicting I and ν cue values was small. The same noise-burst and speech signals used in the previous baseline experiment (Sec. III) were used as stimuli in the present experiments.

3. Procedure

The procedure was identical to that described for the previous apparent source distance experiment (Sec. III), except that the number of blocks of trials completed for each stimulus condition was increased. The first condition used the noise-burst stimulus and source angles of either 0° or 90° azimuth. Source azimuth was the same within a block of 120 trials (10 repetitions per source distance). Both I and ν were independently perturbed from trial to trial, with $\sigma_I = \sigma_\nu = 3$ dB. A total of eight blocks of trials was collected from each listener, four blocks for each source angle. The second condition proceeded identically, except that the speech syllable stimulus was used. The third verification condition also

proceeded in the same way, using the noise-burst stimulus with $\sigma_I = 3$ dB and $\sigma_\nu = 6$ dB. Approximately eight 2-h sessions were required for each listener to complete all three experiments.

B. Weighting procedure

The general weighting procedure used here was very similar to that proposed by Lutfi (1995), with one major difference. The current experiments employed a continuous response type, whereas Lutfi (Lutfi, 1995; Lutfi and Oh, 1997) as well as others (Ahumada and Lovell, 1971; Berg, 1989) have all applied weighting analyses to binary response data. In practice, this detail had little effect on the way the analysis was performed, which was simply a multiple regression analysis on standardized scores.

From the previous apparent distance experiments, it was shown that a power function relates perceived distance to physical sound source distance. This power function was equivalently expressed as a linear function relating logarithmically transformed perceived distances to physical distances (as in Figs. 8–11). This linear function may be expressed in terms of the following simple regression equation:

$$response' = B_r X_r + C + \varepsilon,$$

where $response'$ represents the predicted log-transformed responses, and X_r , the log-transformed physical distances. B_r is the function's slope parameter, C is a logarithmic constant, and ε an error term. For the current experiments, additional terms for explanation of response variance from the independent perturbations applied to I and ν (in units of decibels) were required. Certain interaction terms for I and ν cues with source distance (log-transformed) were also included, in order to evaluate changes in these cue weights with source distance. Predicted responses may now be characterized by the following multiple regression equation:

$$response' = B_r X_r + B_I X_I + B_\nu X_\nu + B_{I \times r} X_I X_r + B_{\nu \times r} X_\nu X_r + C + \varepsilon.$$

The interaction terms of $I \times \nu$ and $I \times \nu \times r$ were not included in this weighting equation, because in subsequent analyses, these terms were found to produce statistically non-significant weights in all cases. This result was expected given the independence of the I and ν perturbations.

Since the predictor variables were not all in the same units and do not all have the same variance, it is difficult to directly interpret the values of B_i in terms of perceptual weights. For these reasons, all variables were standardized (converted to z scores) prior to the multiple regression. Standardized predicted responses are therefore given by the relation

$$Z_{response'} = \beta_r Z_r + \beta_I Z_I + \beta_\nu Z_\nu + \beta_{I \times r} Z_I Z_r + \beta_{\nu \times r} Z_\nu Z_r + \varepsilon.$$

This is the weighting equation utilized for all analyses in this section, where the β_i weights are all partial correlations between the given term (either log-distance, or cue perturbation value, or interaction) and the log-transformed responses. Partial correlation, which conceptually is the correlation be-

tween a given predictor variable and the dependent variable while holding contributions of all other predictor variables constant, is required in this application, because the interaction terms are not independent of the other predictor variables (Lutfi, 1995).

Weights, β_i , were determined for each listener individually, with R^2 , the proportion of total variance accounted for, being used as a measure of how well the weighting equation fits the data. The desired number of trials per condition was determined through a power analysis, with 480 trials allowing for detection of $\beta_i = 0.1$ with 90% power and $\alpha = 0.05$, assuming a population $R^2 = 0.6$ (Cohen and Cohen, 1983).

In an attempt to more precisely quantify the effects of the I and ν cue perturbations on response accuracy, simple power-function fits to the data were also computed as described in Sec. III. These fits were then compared (within listeners) to the baseline conditions (Sec. III) in which no artificial perturbation was applied to any of the distance cues. The fitted exponents were expected to be slightly smaller than the baseline values, because the additional I and ν cue perturbations added some degree of uncertainty to the stimulus and this uncertainty likely causes a slight regression towards a mean perceived distance.

C. Results

Figures 12 and 13 display the individual listener weights for the noise and speech stimulus conditions ($\sigma_I = \sigma_\nu = 3$ dB), respectively, each with separate analyses for 0° [Figs. 12(a) and 13(a)] and 90° degrees [Figs. 12(b) and 13(b)]. A number of observations apply to all results shown in Figs. 12 and 13. First, it is observed that r' weights are always very large and positive. This result is expected given the relatively small amounts of experimentally imposed perturbation driving the additional weights. Second, I and ν weights are always negative. This is a result of the inverse relation between the I and ν cues and source distance. Third, all weighting equations represent good fits to the response data, explaining approximately 70% of the total variance, R^2 .

Considerably more interesting is the way weighting strategies appear to change for different source signals or source directions. To evaluate these changes fully, multiple repeated-measures ANOVAs were performed treating each correlation weight as a dependent variable. Because the sampling distribution of nonzero correlations is skewed, however, weights were transformed via Fisher's z' transform (Cohen and Cohen, 1983)

$$z' = \frac{1}{2}[\log_e(1 + \beta_i) - \log_e(1 - \beta_i)],$$

prior to the ANOVAs. In an attempt to hold experiment-wise $\alpha = 0.05$, α was split between the dependent variable analyses in the following way. Since I and ν perturbation values were chosen independently, each was evaluated at a nonsplit level of $\alpha = 0.05$. The remaining dependent variables were each evaluated at the $\alpha = 0.05/3 = 0.0167$ level. Both angular source position (0° or 90°) and stimulus type (noise or speech) were treated as within-subjects factors for all analyses.

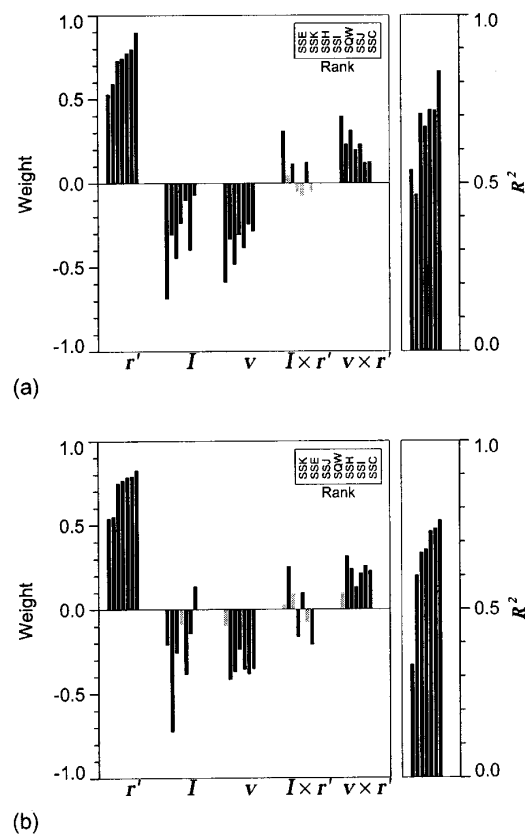


FIG. 12. Individual listener weights for the noise-burst stimulus ($\sigma_I = \sigma_\nu = 3$ dB), with the 0° azimuth results displayed in the top panel (a), and 90° azimuth results in the bottom panel (b). The order of individual listener weights from left to right is displayed in the upper right corner of each panel. This rank from left to right is determined by the r' weight from lowest to highest. Each bar corresponds to the weight for one listener and one cue. Lightly shaded bars represent weights that are nonsignificantly ($p > 0.05$) different from zero. The proportion of variance explained by the weighting model for each listener, R^2 , is displayed on the right side of both panels.

A number of differences in the weights between experimental conditions may be seen. First, the ν weight changes significantly between noise and speech stimulus types, $F(1,6) = 24.478$, $p = 0.003$ ($\eta^2 = 0.803$), with greater weight (in absolute value) being placed on the ν cue for the noise-burst stimulus (Fig. 12) than for the speech stimulus (Fig. 13). Second, the $\nu \times r'$ weight also changes significantly between the stimulus types, $F(1,6) = 13.010$, $p = 0.011$ ($\eta^2 = 0.684$). For the noise-burst stimulus (Fig. 12), nearly all $\nu \times r'$ weights were significantly different from zero. By rearranging terms in the raw-score weighting equation

$$Z_{\text{response}'} = B_{r'}X_{r'} + (B_I + B_{I \times r'}X_{r'})X_I + (B_\nu + B_{\nu \times r'}X_{r'})X_\nu + C + \varepsilon,$$

it is clear that significant $\nu \times r'$ weight implies that the ν weight increases (in absolute value) with increasing source distance. This result was not observed for the speech stimulus (Fig. 13), which produced very few significant $\nu \times r'$ weights. Third, I weights are significantly different for

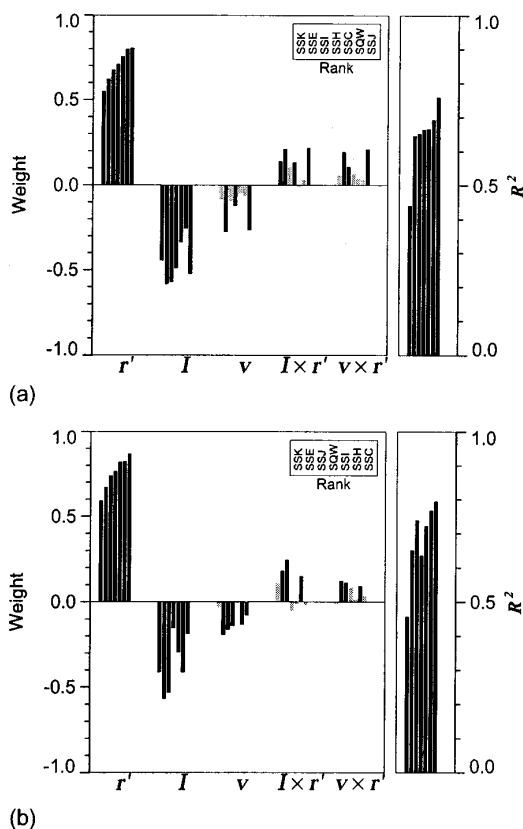


FIG. 13. Individual listener weights for the speech stimulus ($\sigma_I = \sigma_v = 3$ dB), with the 0° azimuth results displayed in the top panel (a), and 90° azimuth results in the bottom panel (b). The order of individual listener weights from left to right is displayed in the upper right corner of each panel. This rank from left to right is determined by the r' weight from lowest to highest. Each bar corresponds to the weight for one listener and one cue. Lightly shaded bars represent weights that are nonsignificantly ($p > 0.05$) different from zero. The proportion of variance explained by the weighting model for each listener, R^2 , is displayed on the right side of both panels.

source angles of 0° and 90° azimuth, $F(1,6) = 12.423$, $p = 0.012$ ($\eta^2 = 0.674$), with I weight generally higher for 0° [Figs. 12(a) and 13(a)] and lower for 90° [Figs. 12(b) and 13(b)].

Since simple power-function fits to the data were also performed, these results may be compared, within the same listeners, to those described in Sec. III. Specifically, the slopes relating log-transformed perceived distance to log-transformed physical distance were compared (which are equivalent to power-function exponents). As predicted, slopes were shallower in the weighting experiments than for the unperturbed baseline conditions, $t(29) = 3.752$, $p = 0.001$. The slope decrement was only approximately 20% on average, however. Hence, it is concluded that the naturalistic stimulus conditions of the previous experiments have not been unduly sacrificed in the name of experimental cue manipulation and experimental design. It is also important to note that, just as in the baseline experiments, slopes did not change significantly within listeners when either source type,

$F(1,6) = 1.053$, $p = 0.344$, or source orientation was changed, $F(1,6) = 1.414$, $p = 0.279$.

Results of the third control experiment, in which $\sigma_I = 3$ dB and $\sigma_v = 6$ dB, verified that precise choices of σ do not affect the resulting weighing patterns. Multiple repeated-measures ANOVAs were used to compare the results of the third control experiment to the first experiment. No statistically significant differences were found. Because of the standardization of scores, the differences in σ do not affect the weighting procedure itself. This result demonstrates, however, that a listener's responses are also unaffected by differences in perturbation σ . Hence, within a reasonable range, the precise choices for σ_I and σ_v are probably irrelevant.

In summary, the perceptual weights assigned to intensity and direct-to-reverberant distance cues were found to change substantially as a function of source signal type, source direction, and to a lesser extent, source distance. The ecological validity of this result is strengthened by the fact that the weights were determined under conditions of largely natural combinations of cues (i.e., small cue perturbation values).

V. GENERAL DISCUSSION

The results of the experiments reported in this article demonstrate that: (1) listener's judgments of apparent sound source distance produce consistent patterns of error under a variety of stimulus conditions, including different source signals (speech or noise) and different source directions (0° or 90° azimuth). (2) The manner in which listeners process, or perceptually weight two principal distance cues (intensity and direct-to-reverberant ratio) varies substantially across the same stimulus conditions. These results taken together point to an auditory distance perception process that, though adept at modifying its cue integration rules in response to varying stimulus conditions, consistently produces distance percepts that are in error when compared to physical states of affairs. At least two questions remain, however. (1) How do the errors reported in the current experiments relate to errors that have been reported in past studies? (2) How might the perceptual weighting process shown here generalize to other stimulus and environmental conditions?

A. Apparent distance: Relation to past results

Almost without exception, past studies of auditory distance perception have reported that listeners tend to underestimate the true physical distance to a source of sound. Although the results of the present experiments are consistent with this general statement, it is of interest to more closely compare the nature and degree of underestimation with relevant past results. To do this, past studies were identified in which more than two far-field source distances were used and listeners made judgments of apparent distance. Table I summarizes the results of all known past experiments that satisfy these two criteria. For each study, the reported data were fit with power functions. Parameters of the fitted functions are also displayed in Table I.

High values of R^2 were obtained in all cases, indicating that power functions represent good fits to the data. The average exponent, a , from these studies was 0.59, with s.d. = 0.24, and average constant, k , was 1.66 (s.d. = 0.92). This

TABLE I. Summary of results from past auditory distance perception studies in which more than two source distances were presented, and listeners made judgments of apparent distance. Each study is described by environment in which the experiment took place, source stimulus type, source azimuth angle (θ), as well as range and number (n) of distances presented. Power functions of the form $\psi_r = k \phi_r^a$ (where ψ_r is perceived distance and ϕ_r is the physical source distance) were fit to the published data. Parameters of the fitted functions (a and k), along with R^2 are displayed for each study.

	ID	Environment	Stimulus	θ	Range (m)	n	a	k	R^2	Notes
Békésy (1949)	1	Anechoic	Live speech	0°	1–10	7	0.76	1.02	0.94	
Coleman (1968)	2	Damped room	LP filtered (10.56 kHz) impulse	0°	2.44–7.92	10	0.38	3.11	0.94	
	3	Damped room	LP filtered (7.68 kHz) impulse	0°	2.44–7.92	10	0.39	3.38	0.93	
Holt and Thurlow (1969)	4	Outdoors	Noise burst ^a	90°	9.75–19.2	4	0.75	1.46	0.97	
Mershon and Bowers (1979)	5	Room	Noise ^a	0°	0.55–8	5	0.65	1.38	0.87	b
	6	Room	Noise ^a	90°	0.55–8	5	0.65	1.46	0.89	b
	7	Room	Noise ^a	0°	0.55–8	5	0.59	1.40	0.89	c
	8	Room	Noise ^a	90°	0.55–8	5	0.74	1.44	0.94	c
Mershon <i>et al.</i> (1989)	9	Reverberant room	Pulse train	0°	0.75–6	4	0.67	2.63	0.61	
	10	Reverberant room	Pulse train	Variable	0.75–6	4	0.65	1.88	0.94	
	11	Damped room	Pulse train	0°	0.75–6	4	0.47	0.40	0.66	
	12	Damped room	Pulse train	Variable	0.75–6	4	0.68	0.39	0.59	
	13	Reverberant room	Pulse train	Variable	0.75–6	4	0.90	1.76	0.62	d
	14	Reverberant room	Pulse train	Variable	0.75–6	4	0.55	2.64	0.88	e
	15	Damped room	Pulse train	Variable	0.75–6	4	0.77	0.24	0.51	d
	16	Damped room	Pulse train	Variable	0.75–6	4	0.38	0.66	0.29	e
Nielsen (1991)	17	Listening room	Speech ^a	0°	1–5	4	0.57	1.78	0.95	f
	18	Anechoic	Speech ^a	0°	1–5	4	−0.12	1.48	0.54	f
	19	Classroom	Speech ^a	0°	1–5	4	0.59	1.70	0.96	f
	20	Listening room	Speech ^a	−45°	1–2.82	3	0.36	2.15	0.97	f
	21	Anechoic	Speech ^a	45°	1–5	4	0.05	1.60	0.96	f
	22	Classroom	Speech ^a	45°	1–5	4	0.55	1.85	0.98	f
	23	Listening room	Speech	0°	1–5	4	0.76	1.38	0.98	
	24	Listening room	Guitar	0°	1–5	4	0.70	1.11	0.98	
	25	Listening room	Pop music	0°	1–5	4	0.95	1.04	0.95	
	26	Listening room	Noise	0°	1–5	4	0.94	0.93	0.95	
Speigle and Loomis (1993)	27	Outdoors	Pulse train	0° to 90°	2–6	3	0.29	3.14	0.96	g
Ashmead, Davis and Northington (1995)	28	Outdoors	Noise burst (random intensity)	0°	5–19	15	0.50	4.04	0.76	h
	29	Outdoors	Noise burst (random intensity)	0°	5–19	8	0.60	3.04	0.97	i
Loomis <i>et al.</i> (1998)	30	Outdoors	Synthesized speech	0°	4–16	4	0.78	1.02	0.99	
	31	Outdoors	Synthesized speech	−120° to 120°	4–16	3	0.79	0.83	>0.99	g
Bronkhorst and Houtgast (1999)	32	Damped room	Noise burst ^a	0°	1–3	5	0.86	0.90	0.97	
	33	Reverberant room	Noise burst ^a	0°	1–3	5	0.44	1.66	0.98	

^aConstant at-the-ear intensity.

^bFirst presentation.

^cSecond presentation.

^dLow S/N.

^eHigh S/N.

^fAveraged across source level.

^gAveraged across θ .

^hExperiment 1.

ⁱExperiment 2.

result suggests that, although the average exponent observed in past studies is generally higher than those observed here, interexperiment exponent variability is quite large. As a result, both the exponents and constants observed in the current experiments may arguably be considered similar to those observed in the past, because average distance exponent and constant values observed here are within 1 standard deviation of the past studies. Of additional note, the only experiments listed in Table I to use virtual acoustic methods were those of Bronkhorst and Houtgast (1999). One of these experiments, conducted under reverberant conditions (ID 33), produced an exponent value of 0.44, which is very similar to the average exponent of 0.39 observed in the current experiments.

These results showing that far physical distances are underestimated and near distances are overestimated are consistent with an influence of a perceptual organization factor known as *specific distance tendency* (Gogel, 1969; Mershon

and King, 1975). This factor describes a tendency towards a specific “default” value of perceived distance under conditions where all distance cues are removed. An estimate of this distance tendency may be computed by determining the point at which perceived distance is equivalent to physical distance. Assuming that a power function does closely approximate the underlying distance psychophysical function, and using the average exponent and constant resulting from the current experiments, the value of specific distance tendency is estimated to be approximately 1.5 m. This value is in good agreement with average source distance estimates that have been reported under anechoic conditions using a noise-burst sound source with fixed level, conditions where distance is in theory indeterminate (Wightman and Kistler, 1997; Zahorik, Kistler, and Wightman, 1994).

To evaluate the effects of differing stimulus conditions on perceived distance, the past experiments shown in Table I were grouped according to environmental factors (either re-

reverberant or nonreverberant), source stimulus factors (either familiar or unfamiliar), and source angular positions (either median plane or nonmedian plane). Environments listed in Table I as “room,” “reverberant room,” “listening room,” or “classroom” were classified as reverberant. All other environments were nonreverberant. Sound source types listed as “live speech,” “speech,” “synthesized speech,” “guitar,” or “pop music” were classified as familiar. All other source types were labeled as unfamiliar. Finally, source azimuth angles listed as 0° were classified as median-plane positions (no 180° positions existed), and all other positions were considered nonmedian plane. Differences between these groups for both exponent, a , and constant, k , values were examined via two full-factorial ANOVAs. These analyses revealed no statistically significant differences between any of the stimulus condition groups, or their interactions for either dependent variable.

When results from additional distance perception studies that did not meet the selection criteria for Table I are also considered, some disagreement may be noted as to the effects of source stimulus familiarity as well as source angular position on distance perception accuracy. Both Coleman (1962) and McGregor *et al.* (1985) report increased distance judgment accuracy for familiar sources. These results are at odds with those of Nielsen (1991) that show no significant differences in apparent distance judgments between four types of stimuli, both familiar and unfamiliar (see ID numbers 23–26 in Table I). Similarly, results from certain studies show that judgment error is related to source direction for both near-field (Brungart, Durlach, and Rabinowitz, 1999) and far-field sources (Gardner, 1969; Holt and Thurlow, 1969; Mershon *et al.*, 1989). Results from other studies do not show this relationship in either the near-field (Simpson and Stanton, 1973) or the far-field (Cochran, Throop, and Simpson, 1968; Mershon and Bowers, 1979; Nielsen, 1993).

B. Acoustic cue combination

Perhaps implicit in the cue-weighting results reported in this article is a general framework for combining information from multiple cues in order to arrive at a unitary distance percept. Consider a framework in which each of the individual distance cues available to the listener produces its own distance estimate, estimates which are then combined and weighted according to the “quality” of distance information they provide for the given situation. Quality is determined by two factors: content of the auditory scene, and consistency of the distance estimate for a given cue with those estimates from other cues. As a result, not only are cues that are unavailable or unreliable in a given scene weighted less, but cues that produce distance estimates that are highly discrepant from the other cue estimates are also down-weighted. This general framework for cue combination is essentially identical to that proposed by Landy *et al.* (1995) for visual depth perception.

In the current experiments, such a cue combination framework could perhaps explain the changes in the perceptual weights placed on different acoustic distance cues. For example, the intensity cue was found to be weighted more heavily than the direct-to-reverberant cue for the speech sig-

nal (Fig. 13). This implies, based on the hypothesized cue combination framework, that intensity is a more reliable cue than direct-to-reverberant ratio in this situation. This implication seems plausible, given that speech is a familiar signal, and prior knowledge of natural speech levels may be used to decrease and possibly eliminate ambiguity present in the intensity cue (Brungart and Scott, 2001). In contrast, direct-to-reverberant ratio depends on potentially more variable and complicated acoustic information about the environment. For unfamiliar signals, the intensity cue may become less reliable than direct-to-reverberant energy ratio, given the lack of natural level information available. The patterns of weights for the noise-burst signal (Fig. 12) are consistent with this notion.

This cue combination framework could also accommodate distance perception in varying acoustic environments or over differing source ranges. For the acoustic environment in which the current experiments were conducted, intensity and direct-to-reverberant ratio were the primary cues available to listeners. Other environments and situations may warrant the use of different cues. For example, when sources are very far away, changes in high-frequency energy with distance become increasingly large, and therefore likely receive significant perceptual weight in these circumstances. Similarly, very close distances in anechoic space produce substantial changes in the binaural cues as distance is varied, and therefore also likely receive significant weight. These examples of the extremely variable nature of distance information present in the acoustic stimulus underscore the importance of flexibility in combining information from various acoustic distance cues to produce distance percepts that are unitary and stable, although often underestimates of physical source distance.

ACKNOWLEDGMENTS

I would like to thank Fred Wightman and Doris Kistler for their invaluable discussions and suggestions, Kathy Hoopes for her assistance in behavioral data collection, and Bob Lutfi, Rick Jenison, Jim Dannemiller, and John Brugge for providing helpful comments on an earlier version of this work. Financial support was provided by the NIH-NIDCD Nos. (P01 DC00116-23), ONR (N00014096-1-0675), NASA (Cooperative Agreement No. NCC2-542), and NIH-NEI (No. 1F32EY07010-01).

- Ahumada, A., and Lovell, J. (1971). “Stimulus features in signal detection,” *J. Acoust. Soc. Am.* **49**, 1751–1756.
- Ashmead, D. H., Davis, D. L., and Northington, A. (1995). “Contribution of listeners’ approaching motion to auditory distance perception,” *J. Exp. Psychol. Hum. Percept. Perform.* **21**, 239–256.
- Békésy, G. v. (1949). “The moon illusion and similar auditory phenomena,” *Am. J. Psychol.* **62**, 540–552.
- Berg, B. G. (1989). “Analysis of weights in multiple observation tasks,” *J. Acoust. Soc. Am.* **86**, 1743–1746.
- Berg, B. G. (1990). “Observer efficiency and weights in a multiple observation task,” *J. Acoust. Soc. Am.* **88**, 149–158.
- Berg, B. G., and Green, D. M. (1990). “Spectral weights in profile listening,” *J. Acoust. Soc. Am.* **88**, 758–766.
- Blauert, J. (1983). *Spatial Hearing* (MIT Press, Cambridge, MA), p. 118.
- Bronkhorst, A. W., and Houtgast, T. (1999). “Auditory distance perception in rooms,” *Nature (London)* **397**, 517–520.

- Brungart, D. S., Durlach, N. I., and Rabinowitz, W. M. (1999). "Auditory localization of nearby sources. II. Localization of a broadband source," *J. Acoust. Soc. Am.* **106**, 1956–1968.
- Brungart, D. S., and Rabinowitz, W. M. (1999). "Auditory localization of nearby sources. Head-related transfer functions," *J. Acoust. Soc. Am.* **106**, 1465–1479.
- Brungart, D. S., and Scott, K. R. (2001). "The effects of production and presentation level on the auditory distance perception of speech," *J. Acoust. Soc. Am.* **110**, 425–440.
- Brunswick, E. (1952). *The Conceptual Framework of Psychology* (University of Chicago Press, Chicago).
- Cochran, P., Throop, J., and Simpson, W. E. (1968). "Estimation of distance of a source of sound," *Am. J. Psychol.* **81**, 198–206.
- Cohen, J., and Cohen, P. (1983). *Applied Multiple Regression/Correlation Analysis for the Behavioral Sciences* (Erlbaum, Hillsdale, NJ).
- Coleman, P. D. (1962). "Failure to localize the source distance of an unfamiliar sound," *J. Acoust. Soc. Am.* **34**, 345–346.
- Coleman, P. D. (1968). "Dual role of frequency spectrum in determination of auditory distance," *J. Acoust. Soc. Am.* **44**, 631–634.
- Da Silva, J. A. (1985). "Scales for perceived egocentric distance in a large open field: Comparison of three psychophysical methods," *Am. J. Psychol.* **98**, 119–144.
- Duda, R. O., and Martens, W. L. (1998). "Range dependence of the response of a spherical head model," *J. Acoust. Soc. Am.* **104**, 3048–3058.
- Foster, S. (1986). "Impulse response measurements using Golay codes," in *IEEE, 1986, Conference on Acoustics, Speech, and Signal Processing (ICASSP)* (IEEE, New York), Vol. 2, pp. 929–932.
- Gardner, M. B. (1969). "Distance estimation of 0 degrees or apparent 0 degree-oriented speech signals in anechoic space," *J. Acoust. Soc. Am.* **45**, 47–53.
- Gogel, W. C. (1969). "The sensing of retinal size," *Vision Res.* **9**, 1079–1094.
- Green, D. M., and Mason, C. R. (1985). "Auditory profile analysis: Frequency, phase, and Weber's Law," *J. Acoust. Soc. Am.* **77**, 1155–1161.
- Hartley, R. V. L., and Fry, T. C. (1921). "The binaural location of pure tones," *Phys. Rev.* **18**, 431–442.
- Holt, R. E., and Thurlow, W. R. (1969). "Subject orientation and judgment of distance of a sound source," *J. Acoust. Soc. Am.* **46**, 1584–1585.
- Ingard, U. (1953). "A review of the influence of meteorological conditions on sound propagation," *J. Acoust. Soc. Am.* **25**, 405–411.
- Jesteadt, W., Wier, C. C., and Green, D. M. (1977). "Intensity discrimination as a function of frequency and sensation level," *J. Acoust. Soc. Am.* **61**, 169–177.
- Landy, M. S., Maloney, L. T., Johnston, E. B., and Young, M. (1995). "Measurement and modeling of depth cue combination: In defense of weak fusion," *Vision Res.* **35**, 389–412.
- Loomis, J. M., Klatzky, R. L., Philbeck, J. W., and Golledge, R. G. (1998). "Assessing auditory distance perception using perceptually directed action," *Percept. Psychophys.* **60**, 966–980.
- Lutfi, R. A. (1995). "Correlation coefficients and correlation ratios as estimates of observer weights in multiple-observation tasks," *J. Acoust. Soc. Am.* **97**, 1333–1334.
- Lutfi, R. A., and Oh, E. L. (1997). "Auditory discrimination of material changes in a struck-clamped bar," *J. Acoust. Soc. Am.* **102**, 3647–3656.
- McGregor, P., Horn, A. G., and Todd, M. A. (1985). "Are familiar sounds ranged more accurately?" *Percept. Mot. Skills* **61**, 1082.
- Mershon, D. H., Ballenger, W. L., Little, A. D., McMurtry, P. L., and Buchanan, J. L. (1989). "Effects of room reflectance and background noise on perceived auditory distance," *Perception* **18**, 403–416.
- Mershon, D. H., and Bowers, J. N. (1979). "Absolute and relative cues for the auditory perception of egocentric distance," *Perception* **8**, 311–322.
- Mershon, D. H., and King, E. (1975). "Intensity and reverberation as factors in the auditory perception of egocentric distance," *Percept. Psychophys.* **18**, 409–415.
- Møller, H., Sørensen, M. F., Hammershøi, D., and Jensen, C. B. (1995). "Head-related transfer functions of human subjects," *J. Audio Eng. Soc.* **43**, 300–321.
- Nielsen, J. L. (1998). "Improvement of signal-to-noise ratio in long-term MLS measurements with high-level nonstationary disturbances," *J. Audio Eng. Soc.* **45**, 1063–1066.
- Nielsen, S. H. (1991). "Distance perception in hearing," Ph.D. thesis, Aalborg University, Denmark.
- Nielsen, S. H. (1993). "Auditory distance perception in different rooms," *J. Audio Eng. Soc.* **41**, 755–770.
- Pearson, E. S. (1931). "The analysis of variance in cases of non-normal variation," *Biometrika* **23**, 113–131.
- Richards, D. G., and Wiley, R. H. (1980). "Reverberations and amplitude fluctuations in the propagations of sound in a forest: Implications for animal communication," *Am. Nat.* **115**, 381–399.
- Rife, D. D., and Vanderkooy, J. (1989). "Transfer-function measurement with maximum-length sequences," *J. Audio Eng. Soc.* **37**, 419–444.
- Schroeder, M. R. (1965). "New method of measuring reverberation time," *J. Acoust. Soc. Am.* **37**, 409–412.
- Schroeder, M. R. (1979). "Integrated-impulse method measuring sound decay without impulses," *J. Acoust. Soc. Am.* **66**, 497–500.
- Simpson, W. E., and Stanton, L. D. (1973). "Head movement does not facilitate perception of the distance of a source of sound," *Am. J. Psychol.* **86**, 151–159.
- Speigle, J. M., and Loomis, J. M. (1993). "Auditory distance perception by translating observers," in *Proceedings of the IEEE Symposium on Research Frontiers in Virtual Reality* (IEEE, New York), pp. 92–99.
- Stellmack, M. A., Willihnganz, M. S., Wightman, F. L., and Lutfi, R. A. (1997). "Spectral weights in level discrimination by preschool children: Analytic listening conditions," *J. Acoust. Soc. Am.* **101**, 2811–2821.
- Wightman, F. L., and Kistler, D. J. (1989). "Headphone simulation of free-field listening. I. Stimulus synthesis," *J. Acoust. Soc. Am.* **85**, 858–867.
- Wightman, F. L., and Kistler, D. J. (1997). "Monaural sound localization revisited," *J. Acoust. Soc. Am.* **101**, 1050–1063.
- Willihnganz, M. S., Stellmack, M. A., Lutfi, R. A., and Wightman, F. L. (1997). "Spectral weights in level discrimination by preschool children: Synthetic listening conditions," *J. Acoust. Soc. Am.* **101**, 2803–2810.
- Zahorik, P. (2000). "Limitations in using Golay codes for head-related transfer function measurement," *J. Acoust. Soc. Am.* **107**, 1793–1796.
- Zahorik, P., Kistler, D. J., and Wightman, F. L. (1994). "Sound localization in varying virtual acoustic environments," in *Proceedings of the Second International Conference on Auditory Display, ICAD 1994*, Santa Fe, NM, edited by G. Kramer, pp. 179–186.
- Zhou, B., and Green, D. M. (1992). "Characterization of external ear impulse responses using Golay codes," *J. Acoust. Soc. Am.* **92**, 1169–1171.