PSYCHOPHYSICAL CALIBRATION FOR CONTROLLING THE RANGE OF A VIRTUAL SOUND SOURCE: MULTIDIMENSIONAL COMPLEXITY IN SPATIAL AUDITORY DISPLAY

William L. Martens

Multimedia Systems Lab.
University of Aizu
Aizu-Wakamatsu 965-8580, Japan
wlm@u-aizu.ac.jp

ABSTRACT

Just as control over perceived azimuth and elevation of a virtual sound source should be psychophysically calibrated in spatial auditory display, so should perceived range; however, in contrast to azimuth and elevation display, precise control over auditory range has been difficult to achieve. This is partly due to the multidimensional complexity of the human response to spatial auditory stimulation, but it is also due to the multidimensional complexity of the acoustic stimulus for range, which includes a substantial number of independent parameters even in the case of static spatial positioning of sound source relative to listener. In the static case, there is strong dependence of perceived range upon at least the following display parameters: direct sound level, indirect sound level, interaural cross-correlation, and the relation between direct and indirect sound spectra associated with air absorption and close-range head-related effects. If the sound source range varies smoothly over time, other display parameters (such as dynamic variation in pitch of the direct sound, or Doppler shift, and also dynamic variation in the initial time gap) become significant, and interact with the above-listed parameters to produce changes in auditory range that have proven difficult to successfully model. In the absence of a model that integrates variation in all of these display parameters and successfully predicts range variation, two reasonable solutions to the problem of range control present themselves. The first is to base control upon highly realistic simulation, relying on the relatively good match between perceived range and specified range that can be observed when nearly all displayed auditory spatial information is consistent with an adequate physical model. The second solution is to base control upon psychophysical range judgments under conditions of expected use of the display, relying on an inversion of a range prediction model fit to the judgments using multiple regression analysis. This paper presents two examples of successful psychophysical calibration for auditory range control for spatially static sources: One case employed a simplified model of range-dependence in simulated head-related transfer functions for headphone display of virtual sources at close range (within the listener's personal space). The other case employed a room-related (rather than head-related) loudspeaker-based sound simulation to create auditory imagery of sources relatively far away from a group of simultaneous listeners. This room-related loudspeaker system was designed with the pragmatic goal of reducing reliance upon fixed, known listening locations. In both cases, adequate control over the range of a set of sound sources (short speech samples) was achieved using a look-up table derived by inverting the range prediction equation fit to collected human range ratings.

1. INTRODUCTION

1.1. Psychophysical validation versus calibration

The development of technology for the auditory display of virtual sound sources in virtual acoustical spaces has received great benefits from a long history of research on the acoustic cues used by human listeners in the formation of auditory spatial imagery. These benefits have accrued primarily from the results of psychophysical investigations, controlled studies of the relationship between observed physical features of the sound stimuli, and reported attributes of the perceptual response to those stimuli. In contrast to such basic scientific research, this paper emphasizes a more applied approach to psychophysical investigation. Whereas the primary motivation in spatial hearing research has been to gain greater understanding of the mechanisms of human spatial hearing, the motivation for this applied research has been the verification and validation of various spatial audio rendering technologies under development. As the ultimate goal here is the deployment of auditory display technology for which specified responses are in some way calibrated to the actual responses of the human listener, the term "psychophysical calibration" is employed to identify this endeavor.[1]

When auditory and visual displays are integrated into a spatially coordinated human-computer interface, a first step in setting up such a multi-modal display system is to make sure that corresponding points in the respective perceptual spaces are in good registration to each other. The most direct approach to solving this problem is to execute an egocentric cross-modal matching task for the two display modalities. Here, the issue is not so much validation as it is calibration. Validation is what is done when system performance is uncertain at a basic level. Calibration is what is done when a system is known to make distinctions consistently in the displayed attributes for a particular modality, but the question of *mapping* the display space for that sensory modality remains. One premise of this paper is that this sort of calibration should be completed for an auditory display even in the development of spatial interface systems that might be termed "audio-only."

[1] Of course, no mutually exclusive bodies of research exist that might be identified as pure-basic versus pure-applied research. In fact, it might be argued that the interplay between researchers with engineering goals and those with scientific goals provide an essential tension [1] that drives progress for the entire field. The reader is refered to the author's paper [2] on "Uses and misuses of psychophysical methods in the evaluation of spatial sound reproduction" for further discussion of this issue.

It is almost never the case that the perceived locations of virtual sound sources are irrelevant to the deployment of spatial audio rendering technology. Therefore, knowledge of the actual perceived locations will almost always be needed, so that a good mapping between specified and resulting source locations is confirmed. For example, if a virtual space is navigated by a user wearing a head-mounted display (HMD), the perceived locations of invisible virtual acoustic objects must change in an expected way if they are to be perceived as spatially stationary sound sources. In this case, a match is implicitly made between the location (perhaps proprioceptively determined) of the user's ego-center (or self) in physical space and the mental projection of that point into the auditory space created by the virtual acoustic display.

1.2. Veridicality and Constancy

Veridical perception is what is expected when stimulation is adequately full and realistic. Unfortunately, spatial auditory displays typically rely upon incomplete virtual acoustic simulations, and typically provide impoverished range cues (especially the dynamic auditory cues associated with active localization behavior [3]). When veridicality is lost due to overly simplified virtual acoustic simulation, something else is typically lost along with it: *the constancy phenomena*, which are those phenomena that keep variations in source loudness from becoming variations in auditory range as well as variations in source loudness. In order to discuss loudness *constancy*, it is necessary to make clear the distinction between *proximal* and *distal* stimulation:

> "Stimulation defined at the level of the [sense] receptors is called *proximal* stimulation. Stimulation defined at the level of the physical objects outside *O* [(the observer)] is called *distal* stimulation. . . . A proportionality between perceptions and *distal* rather than *proximal* stimulation is termed *the constancy phenomena*." (Gogel [4], p. 367)

For example, if there is *loudness constancy* with varying sound source range, as reported in [5], then the perceived loudness of a source of constant *distal* sound pressure level (SPL) will not change as its physical range is changed (though large changes in *proximal* SPL will be observed).

1.3. Level-Based Range Control: A Fundamental Ambiguity

The typical solution to the problem of controlling auditory range has been to use changing source loudness as the primary cue, but this solution presents an ambiguous stimulus to the user as changes in SPL at the source position are not as easily discriminated as changes in loudness at the listening position. The ambiguity stems from the difficulty of determining whether changes in the *proximal* SPL (i.e., level at the ear) are caused by changes in *distal* SPL (i.e., level at the source position) or by changes in source range. In fact, it is conceivable that a change in level due to a change in source range could be accompanied by a complementary change in level at the source position (*distal* SPL), and these changes would go unnoticed as there would be no net change in level at the ear (*proximal* SPL).

The primary means available for resolving this ambiguity in auditory spatial display of source range is the inclusion of appropriate indirect sound. This factor is especially powerful under conditions in which the synthetic indirect sound is clearly separable from the direct sound, as it is in spatially immersive multi-loudspeaker displays such as the Pioneer Sound Field Controller [6], or PSFC (described more fully in section 3 of this paper). In contrast to conventional two-loudspeaker sound spatialization, sound sources presented via the PSFC remain at relatively stable locations in space as listeners change the position and orientation of their heads. In addition, such loudspeaker arrays avoid most of the difficulties associated with headphone-based presentations using binaural techniques (e.g., convolution with head-related transfer functions, or HRTFs), such as front↔back reversals and problems with externalization. An immersive loudspeaker array enables a direct approach to the creation of enveloping virtual acoustics, resulting in a soundfield stabilization that naturally requires no head-tracking. Of course, as the reverberation simulation is "room-related" rather than "head-related" [7], the system exhibits none of the advantages associated with cross-talk cancellation (*a.k.a.* transaural stereo [8]), which extend to improved display of auditory range as well as direction (e.g., precise control over interaural cross-correlation at the listener's ear is possible only when the system is head-related).

In contrast to the control of auditory range possible in public spaces, there is another means available for resolving level-based ambiguity in the auditory range of virtual sources located at close range. For headphone-based auditory display, a modification of the HRTF deployment can also be implemented to allow for range-based transformation of the direct sound. Such range-based variation in the HRTF has been well documented for sources within arm's reach of the listener [9], and their effectiveness in producing changes in range perception in the listener's "personal space" has been established [10]. Of course, such HRTF-based control is not practicable when using loudspeaker arrays for multiple listeners (such as the "room-based" PSFC system), but it is possible in "head-related" loudspeaker reproduction employing cross-talk cancellation.

1.4. Static Auditory Range Cues

Rather than giving a comprehensive treatment to the psychoacoustics of range, the emphasis here is upon acoustic cues that are readily and often simulated in spatial auditory display. This paper is concerned primarily with how to control auditory range when using simulated cues, and those cues are only briefly summarized here (the reader is refered to [11] for a deeper treatment of acoustic cues used in the control of auditory range.).

The most powerful static auditory range cues in the static case include direct sound level, direct sound spectrum, indirect sound level, and interaural cross-correlation (IACC).[2] In fact, the best

[2]Control of range via manipulation of IACC is more complex than the two cases to be discussed in this paper, but can be very effective especially in loudspeaker reproduction. It is complex because the effect of IACC on range depends on its many parameters (e.g., initial time gap (ITG), frequency band, etc.) that are typically only indirectly controlled through the manipulation of other simulation details. If the gap in time between direct sound and the arrival of the first indirect sound is small enough (as it is in relatively small rooms), then the early indirect sound will be perceptually fused with the direct sound to form a *precedent* image of the sound source that has discriminable depth, breadth, and range. And these perceptual attributes are discriminable from perceptual attributes associated with later-arriving indirect sound [12]. While values of IACC measured over the complete duration of the system response predict auditory source width (an attribute of the *precedent* image), the IACC values measured separately

predictors of auditory range are typically expressed as some combination of these individual parameters. For example, the ratio between direct sound level and indirect sound level, often expressed as the indirect-to-direct (i/d) ratio, is a better predictor of range than direct sound level alone: Though direct sound level is a good relative cue to range, if the direct sound level is decreased in the presence of an indirect sound field of nearly constant level, their interrelation provides a more effective auditory range cue. Indeed, if overall sound level is held constant as the i/d ratio is increased, the reported auditory range of the sound source will also increase [14]. For this reason, in combination with the direct sound level, the i/d ratio often provides additional range control (and predictive power).

In a more complex example, the spectral cues to auditory range are effective mostly via the relation between direct and indirect sound spectra, but the dependence is complex in that the predicted range follows a "U"-shaped function of spectral centroid (a global measure of spectral energy distribution correlated with perceived *brightness*): If all other factors are held constant, then when the spectral centroid of the direct sound is increased relative to that that of the indirect sound, the auditory range of a nearby source will increase, while the auditory range of a distant source will decrease.[3] The physical basis for the confusion is straightforward: Within 2 m of the listener's head, low frequencies are boosted at the ear due to the spherical shape of the wavefront arriving from a nearby source [18] [19], but as range is increased beyond 2 m, the high-frequency components of the nearly planar wavefront lose energy (due to air absorption) more rapidly with increasing range than do the low-frequency components.

1.5. Targeting the Listener's Head or the Listener's Room

In Bauer's seminal 1961 work [20] entitled "Stereophonic earphones and binaural loudspeakers," he describes the means for reproducing head-related auditory spatial imagery via two loudspeakers. Before such a distinction was introduced, perhaps some loudspeaker systems were tacitly room-related, which would have been an appropriate choice for a public auditory spatial display system with more than one simultaneous listener. Though it might be thought that room-related reproduction of indirect sound was anticipated in the early 1960's work on multi-channel spatial sound reproduction at the University of Gottingen (see for example the 1965 paper by Meyer, et al. [21]), the design of that apparatus (which featured 65 loudspeakers!) assumed that the listener was in a fixed location relative to the loudspeakers, and therefore must also be regarded as a head-related spatial auditory display system. It was probably not until Moore's 1983 paper [22] entitled "A general model for spatial processing of sounds," that the concept of a room-related reproduction of indirect sound was explicitly described as a system intended for multiple listeners located at arbitrary locations within the reproduction space. In such room-related spatial sound reproduction, the loudspeaker locations act as windows for sound to enter an imaginary box defined by connecting those loudspeaker locations, and virtual sources are typically localized only outside the borders of this box (localized, that is, outside the boundaries defined by the perceived loudspeaker locations within the listener's auditory space). In contrast, the loudspeakers in head-related spatial sound reproduction should disappear to enable virtual sources to be localized anywhere in the listener's auditory space (nearby or faraway).

These two fundamental types of auditory display can be identified in terms of their assumptions about the spatial relation between listener and display. The reproduction system (headphone- or loudspeaker-based) is identified as "head-related" when the relation between *proximal* stimuli delivered to the user's ears is known and/or determined via either headtracking or immobilization of the user's head. The system is identified as "room-related" if the location of the user's ears relative to sound reproduction devices is disregarded (though perhaps still relevant to system performance). The two systems examined in this paper differ in just these assumptions about the spatial orientation of listener to display. One of the two is a personal display system employing eyescreens and earspeakers (i.e., an HMD); the other is a public display system featuring a large (wide-angle) screen and a multi-loudspeaker array intended for many simultaneous listener – perhaps this case could be termed a room-mounted display (or RMD). The following two sections of this paper deal separately with the perceptual calibration of auditory range display for each of these systems.

2. RANGE CONTROL FOR NEARBY SOURCES: AUDITORY DISPLAY IN PERSONAL SPACE

Typically, personal, headphone-based auditory spatial display technology either fails to project (externalize) the auditory image of the *distal* stimulus to a location in the listener's auditory representation of the surrounding space (i.e., no "out-of-head localization"), or the source may be well externalized, but projected to a location at some greater distance from the listener, most often via the inclusion of a significant amount of reverberation that is easily detectable by the listener. The goal of the experiment reported here, most succinctly put, was to test an efficient means to place a well-externalized virtual sound source so close to the listener's ear that it enters the listener's "personal space" [23]. When the *distally* projected auditory image of a virtual sound source enters the listener's "personal space," a psychological boundary is crossed that potentially carries special meaning to users in particular applications such as teleconferencing in shared virtual acoustic environments. If such an audio transformation were properly engineered (both perceptually valid and perceptually calibrated), a spoken message could be made to sound as if it were whispered into the ear of the recipient, letting them know, for instance, that the message was intended for them in confidence (providing what has been termed a "whisper function" [24]).

One of the continuing problems of headphone-based virtual acoustic imagery has been the difficulty of creating auditory images that are clearly outside of the listener's head using a minimum amount of audio signal processing. The use of HRTFs is conventionally regarded as the first step in creating externalized auditory imagery [25], but this is only moderately effective for some spatial directions of the sound source [26], and truly unreliable [27]. Without simulated indirect sound of some sort, the likely result in headphone listening is "in-head localization" of the auditory im-

for early and later arriving sound are also powerful predictors for many of these perceptual attributes: The late IACC predicts listener envelopment (the sense of feeling spatially surrounded by sound), but the early IACC predicts range of the *precedent* image [13].

[3] In fact, under reduced conditions of observation (i.e., impoverished range cues), an exclusive bias towards one or the other of these two patterns of response has been exhibited by different subjects within a single experiment. Though brighter sources are typically judged closer than darker sources in the free field [15] [16], the addition of unfiltered indirect sound to a set of filtered binaural stimuli caused 6 out of 24 listeners to consistently judge brighter sources to be farther than darker sources [17].

age of the reproduced sound source [28]. The desired result in headphone reproduction, in contrast, is "out-of-head localization" and has often been shown to be dependent upon the amount of reverberation present in sound reproduction (c.f., [29]). But factors other than HRTF-based processing and indirect sound simulation can also aid in externalization. It may be that actively tracking the user's head motion and applying appropriate dynamically varying HRTFs to an input audio signal, can improve externalization. However, the primary benefit of active headtracking in auditory spatial display is the disambiguation of front and rear locations, and the actual benefits in terms of externalization are questionable [27]. The study described below is based upon an unexpected observation of externalized auditory imagery during binaural listening to nearby sound sources in an anechoic chamber. When sources arrive from locations further than 1 m from the listener's head, or near the listener's median plane, externalization is usually not experienced. But well lateralized sources located near the listener's head often do result in externalization. The practical question driving this experiment was that concerning which acoustic cues are most important to include in a simulation intended to produce this close-range externalization.

Of course there are many acoustical sources of information cueing auditory range that are available to the human listener at close range, but this study focuses only on the very simple case of dry HRTF-based processing (i.e., involving no indirect sound simulation). In particular, it was the level-based cues contained in the direct sound that seemed worth examining. The idea that variation of interaural level difference (ILD) in the direct sound might aid the listener in detecting range of nearby sources is not new (the hypothesis was probably first stated clearly by Hartley and Fry in 1922 [30]). Brungart [31] confirmed that listener's indeed do better in localizing nearby anechoic sources when those sources are well lateralized away from the listener's median plane. In the author's related study [10] that compared the relative salience of four sources of acoustical information associated with range perception for the human listener, the level-based features of the direct sound (ipsilateral-ear SPL and ILD) dominated other cues (indirect-to-direct sound ratio and high-frequency attenuation of a filter simulating the head-shadow).

The choice to focus on ILD-based range cues via a simple manipulation of ipsilateral and contralateral SPL was also motivated by the lack of an efficient yet effective technology for creating auditory images that are clearly outside of the listener's head, but nonetheless very close to the listener's ear. Free-field listening in an anechoic chamber confirms that sources within around 1 m of the listener's head are often externalized and have a special quality that tells the listener that the source is nearby. Although measured HRTFs show a rather complex dependence on range, an analysis of the range dependence in the response of an ideal spherical receiver shows one very striking feature that might account for this special quality. That feature is shown in Fig. 1 for a source at 130° azimuth. Though the gain at DC is nearly 0 dB for sources arriving from ranges greater than 2 m, the interaural level difference (ILD) at DC gain grows large as the source approaches the listener's ear. As the gain increases at the listener's ipsilateral ear, the gain decreases at the listener's contralateral ear. In contrast to the relative auditory range cue provided by loudness, it has been hypothesized that this range-dependent variation in ILD might provide a more absolute range cue to the listener for nearby sources, and that these might be effectively externalized without the contribution typically made by indirect sound.

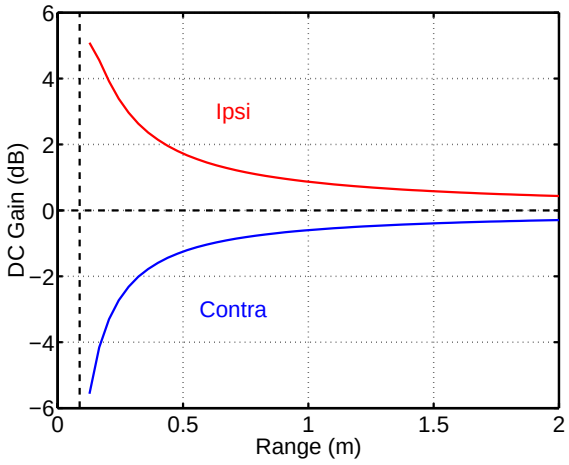

Figure 1: Free-field normalized gain at DC for a point on the surface of an ideal spherical "head" as a function of source range from the center of the sphere. The upper curve (labeled "Ipsi") corresponds to the ipsilateral "ear" response for a source incidence angle of 130° azimuth, while the lower curve (labeled "Contra") corresponds to the contralateral response at that angle. The vertical dashed line shows the model "head" radius of .0875 m. (On the assumption that this paper may be less frequently read in black & white print than in its original electronic form, the graphic includes a color code: the ipsilateral curve is red and the contralateral curve is blue.)

2.1. Stimuli

For this experiment, three short utterances, "ha," "hi," and "hu," whispered by three Japanese talkers, were recorded at a range of 1 meter in a large anechoic chamber at the University of Aizu. Whispered (unvoiced) speech was chosen rather voiced speech, since whispered speech contains relatively less reliable information about actual *distal* SPL. Likely due to the recognition of "vocal effort" associated with particular speech production levels, voiced speech contains range cues that seem to be inherent in the timbre of the sound source itself [32]. The three vowel sounds whispered by the talkers for this study span the vowel space defined by the first two formant frequencies of the vowels, and they represent the extremes of vowel coloration in the Japanese language.[4] The intention of this sound source selection was to allow the spatially-processed stimuli to vary in timbre as widely as possible while maintaining the same aspiration /h/ in the consonant-vowel (CV) stimuli. The transient, high-frequency content of the /h/ consonant was included to provide adequate stimulation for perceptual fusion of the stimulus into a single, coherent auditory image of the whispered speech sound.

The spatial sound processing of the stimuli was a variation of the conventional convolution with a pair of HRTFs (subject MES, see [38]) with no simulated indirect sound. The baseline convo-

[4] On average, the first formant frequencies range from 280 to 750 Hz and the second formant frequencies range from 1100 to 2300 Hz when subjects read word lists [33]. The LPC spectra of the speech stimuli used in this study were calculated for each of the four talkers. The average frequencies of the first two formants for the vowel segments of their recorded speech were approximately the following: /a/ – 750, 1200, /i/ – 280, 2280, and /u/ – 310, 1220 Hz.

lution condition used an anechoically measured HRTF pair for a source arriving from the rear at 130° azimuth and from a range of 1.5 meter. Close range auditory range cues were then manipulated strictly in terms of the relative SPL of the ipsilateral- and contralateral-ear signals. The level of the resulting ipsilateral-ear signal was increased from its baseline level in three 3 dB steps (gain was 0 dB, 3 dB, 6 dB, and 9 dB). The level of the resulting contralateral-ear signal was decreased from its baseline level in three 3 dB steps (gain was 0 dB, -3 dB, -6 dB, and -9 dB). No frequency-dependent component of the head-response model was manipulated in order to match the change in the head shadow as sources approach ranges closer than 1 m (c.f., [10]). A factorial combination of ipsilateral and contralateral level was executed by crossing the four possible values for each stimulus, and each of 9 sound sources (3×3 factorial of talkers and vowels) was combined in a randomized order for these 16 SPL combinations (4×4 factorial of ipsi and contra), to produce the set of 144 stimuli comprising a single experimental listening session (of which three were completed by each subject).

2.2. Subjects

Five subjects voluntarily participated in this experiment, four of whom were students at the University of Aizu. One was the author, a researcher with a substantial history of participation in similar listening tests. All were audiologically normal, with no reported hearing loss.

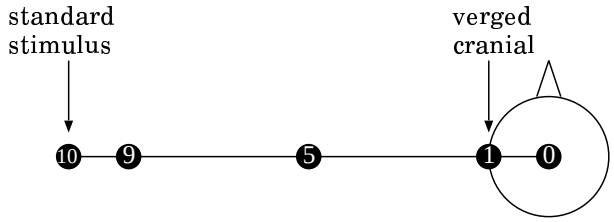

Figure 2: Graphic illustrating the ten-point scale used by listeners to record the range of the comparison stimulus relative to the position of the standard stimulus. See text for details.

2.3. Procedure

The stimuli were presented to the listeners via Sennheiser HD590 headphones using the standard audio conversion hardware of an SGI workstation. The listening test was completed in three blocks of 144 trials each. In each trial, two stimuli were presented with a one-second inter-stimulus interval: a comparison stimulus of variable range, and a standard stimulus of fixed range. The baseline convolution condition (no gain adjustment) served as the standard stimulus that provided a reference by which listeners judged the range of the other experimental stimuli. After hearing first the standard and then the comparison stimulus, the listener was asked to rate the range of the comparison on a scale from 0 to 10. The value of 0 was to be reported if the auditory image was located inside of the listener's head. The value of 1 was to be used if the sound source seemed to be located extremely close to the listener's ear (i.e., the "verged cranial" position). The other extreme of the scale was anchored to the perceived range of the standard stimulus, and the response of 10 was to be given to any source that was perceived at roughly the same range as the that of the standard stimulus. The value of 9 was to be given to a source just noticeably closer than the standard.

A three-minute training session was completed before beginning the experimental trials, during which time listeners were to attempt to establish criteria for their use of the rating scale. The inter-trial interval during the training session was just one second, and this interval was increased to five seconds during experimental trials to allow time for a response to be generated.

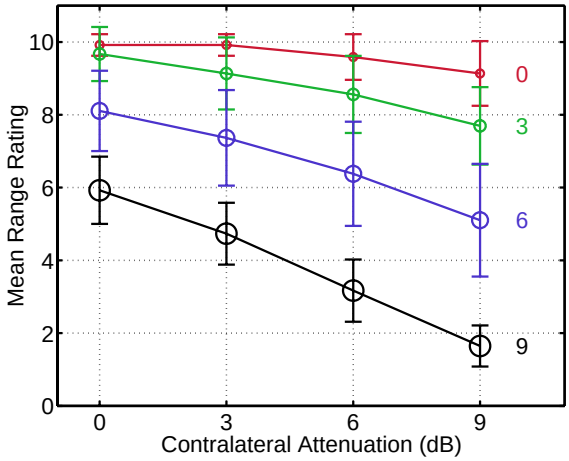

Figure 3: Grand mean range ratings for one subject plotted over contralateral attenuation levels, combining results for three vowel sounds whispered by three native speakers of Japanese. The parameter of the graph is ipsilateral gain level, the values of which are used to label each of the four sets of connected plotting symbols (appearing just to the right of each). In the electronic color version of this paper, the plotting symbols and the connecting lines for the lowest ipsilateral gain level (0 dB) are red (and the plotting symbols for this set are the smallest in the graph). Ratings for the 3 dB gain stimuli are coded green, those for the 6 dB gain stimuli are coded blue, and those for the 9 dB gain stimuli are coded black (and the plotting symbols for this set are the largest in the graph). Standard error bars for each mean follow the same coding scheme.

2.4. Results

The influence of ipsilateral gain and contralateral attenuation on mean range ratings for one subject are shown in Fig. 3. The highest mean range ratings were obtained for a comparison stimulus that was identical to the standard stimulus. Thus, of all the combinations of ipsilateral and contralateral SPL present in the set of comparison stimuli, the greatest source range was reported when the nine sound sources were convolved with the unmodified HRTF (the standard stimulus, and the baseline convolution condition). As the ipsilateral gain was increased, and contralateral attenuation was held constant at 0 dB, reported range decreased for the source, which consistently seemed to be arriving from the rear at around 130° azimuth. The rightmost symbols in the graph shown in Fig. 3 show this decrease from range ratings of around 10 to ratings around 6. The plotting symbols grow larger in the graph as ipsilateral gain is increased, as a reminder that these sources are ap-

proaching the listener (i.e., simulating visual looming). When the ipsilateral gain was held constant at 0 dB, and contralateral attenuation was increased, reported range also decreased, but not to such a great extent. The topmost curve, labeled "0" in the plot, shows decreases from range ratings of around 10 to ratings of only around 9. But when the ipsilateral gain was held constant at an SPL 9 dB greater than that of the standard stimulus, increasing contralateral attenuation produced much lower range ratings (decreasing from ratings of around 6 to ratings below 2). It is clear from these data that larger ILDs result in increasingly closer source localization as the ipsilateral SPL increases. In effect, the close-range ILD cue works best when the whispered speech is so loud that it is already likely to produce a lower range rating.

2.5. Discussion

A general principle that helps to understand these results is that one perception may sometimes depend upon another perception. For example, Gogel [4] made this same point about how visual cues to exocentric distance (i.e., differential displacements) depend upon the observer's judgments of perceived egocentric (absolute) distance:

> "[Because] differential displacements are indeterminate with respect to an exocentric distance without specifying an egocentric distance, a *perception* of egocentric distance is required to translate these cues to a *perception* of exocentric distance." (Gogel [4], p. 367)

It is also true that judgments of auditory range depend not only on perceptual factors, but also upon cognitive factors that influence human perceptual judgment. It is clear that response biases (tendencies) contribute to both absolute (egocentric) and relative (exocentric) distance judgments, and that listeners rely upon these biases more and more as the adequacy of the stimulus cues is reduced. In the absence of strong absolute cues to auditory range, all stimuli will tend to be perceived at some specific range, typically rather near the listener (Gogel [34] termed this tendency the *specific distance tendency*). In static free-field listening, for example, anechoic sound sources produced at a variety of ranges greater than 1 m typically tend to be perceived at a range somewhat less than 1 m, even when the *proximal* SPL is quite low.

The manipulations in the above-described experiment clearly produce relatively strong cues to auditory range, but also demonstrate an interaction between two predictors of range, quantified as ipsilateral gain and contralateral attenuation. It is a straightforward application of Multiple Regression Analysis (MRA) to characterize this interaction in terms of a prediction equation. Regression inversion in the multiple predictor case is well described elsewhere [35]. The inversion of the prediction equation allows a contralateral attenuation value to be determined so that a source at a desired loudness (which specifies the ipsilateral gain value) can also be positioned at a particular range, at least for whispered speech sources arriving at an incidence angle of 130° azimuth. The utility of such inverse prediction may seem limited, but may in fact match well the requirements of some application, such as delivering a voice message from an angle outside of the user's visual field, and at a range close enough to cause the user to notice it immediately. Furthermore, the importance of the spoken message can be indicated by how close to the user's ear the source seems to be located. The details of the inverse prediction via MRA are explained in the context of the following study of faraway sources, since the numerical problem to be solved there is virtually identical to the problem to be solved here (and so a highly redundant explication of the analysis for inverse prediction is here omitted from this paper). Suffice it to say that a multivariate linear equation can be inverted to provide required values of one predictor variable when the values of both the criterion variable and the other predictor variable are specified.

3. RANGE CONTROL FOR FARAWAY SOURCES: AUDITORY DISPLAY IN PUBLIC SPACE

Controlling the perceived direction and range of virtual sound sources is desired in multi-channel loudspeaker-based spatial auditory displays for use in immersive multimedia applications such as teleconferencing and entertainment. The "Synthetic World Zone" at the University of Aizu Multimedia Center is an example of a space designed to present multimodal virtual environments. This immersive multimedia display system presents stereographic visual imagery via three large rear-projection video screens covering a 150° horizontal visual angle (each screen is 3.4 m × 8.1 m). Collocated with this visual display is the Pioneer Sound Field Controller [6], or PSFC, a 15-loudspeaker hemispherical array with a diameter of approximately 10 m. The loudspeakers in this array are positioned in a relatively dry space which comfortably seats about 20 listeners, and together these loudspeakers create a reverberant soundfield simulation with a realistic spatial and temporal distribution of many discrete reflections. The research reported here was focussed upon the performance of this "room-related" spatial auditory display with regard to the localization of virtual sound sources in an immersive virtual acoustic environment; however, it was control of auditory range, in particular, that posed a problem for the PSFC user.

3.1. Pioneer Sound Field Controller

The Pioneer Sound Field Controller (PSFC) is integrated with other display systems in the "Synthetic World Zone" via a personal computer. The system controls the direction of two virtual sound sources via a straightforward interpolation of level between the three loudspeakers surrounding the desired spatial angle (specified by azimuth and elevation angles). The level at each loudspeaker is determined by the angular distance between the virtual source and each loudspeaker (c.f., Gerzon's [36] *Periphony*: "with-height sound reproduction"). Discrete simulated reflections are generated independently for each of the 15 loudspeakers, rather than employing an HRTF to control the direction of each early reflection (as taught by Kendall & Martens [37]). In effect, a room-related transfer function is implemented via transmission through the solid angle subtended by the loudspeaker's "window" that lets sound in from outside of the reproduction space. Thus, the PSFC loudspeakers deliver many simulated discrete reflections, each arriving from the direction of one of the 15 loudspeakers, and each at an appropriate delay and gain, but with no specific listener position assumed (cf. [21]).

In contrast to source direction, control over auditory range of virtual sources using the PSFC is not so straightforward. The loudspeakers are all placed at an approximately constant distance from the room's center, but the range of each virtual source from the listening position is controlled primarily by two parameters: Firstly, the overall volume of the source can be adjusted to create a relative cue to auditory range [39]. Secondly, a more absolute cue to auditory range may be manipulated via the ratio of indirect to direct

Channel	Stimulus Sentence			
1 (standard)	Ford hit raw crime.		No five leave court.	
2 (comparison)		Are raw mouse lush.		Who bought blond fern.

Table 1: Alternating sequential presentation of four short sentences for judging the auditory range of the comparison stimulus relative to the standard stimulus. These are public-domain sources that have been used in the author's previous studies [38].

(i/d) sound levels [40], since the indirect sound level in typical simulated rooms (of fairly large dimensions) is not modulated by source range to as great an extent as the direct sound level is modulated. In fact, the default configuration (which was employed in the current study) is based upon an analysis of the reflection patterns of a large assembly hall (the *Shinjuku Kousei Nenkin Kaikan*), about halfway back from the stage, where indirect sound level remains practically constant when the source range increases from a starting point at least 5 m from the listening position (1.5 m above the floor). Thus as a source recedes from the listener in this space, only the direct sound level decreases, gradually disappearing into the reverberant "grass" presented by the constant-level indirect sound.

The PSFC controls the i/d ratio via a "liveness" parameter (a term perhaps first connected with i/d by Maxfield & Albersheim [41] to describe this psychoacoustic parameter associated with reverberation in enclosed spaces). Though this parameter has also been associated with the term "spaciousness" in related work [42], reverberation time also affects subjective ratings of spaciousness [43] (in German, "Räumlichkeit").[5] In the PSFC system, several parameters are adjusted in an inter-dependent fashion in order to maintain a relatively constant "room impression" (in German, "Raumeindruck"), as changes occur in the liveness or "reverberance" (in German, "Hallichkeit"). The way in which the PSFC handles the complicated situation that arises as liveness is adjusted is the following: First, as the specified liveness is reduced, the rate at which discrete reflection gain falls over time is increased, so that the reverberation time stays fixed (thereby minimizing changes in the perceived size of the simulated room). This change in decay rate of the early reflections allows the level of the last discrete reflection to match the level of the late diffuse reverberant field. Second, the actual PSFC channel volume setting, which is applied to both direct and indirect sound equally, is internally adjusted so as to maintain a constant loudness as the liveness value is manipulated. Since the liveness parameter is usually constant for a given space, progressively changing its value poses something of a dilemma for the listener. As the source loudness remains constant with increasing liveness of the simulated space, the *distal* stimulus seems to be producing progressively greater SPL as it recedes into the distance. This complex situation, complicated even more by the operation of what might truly be termed "loudness constancy" with varying sound source distance [5], underscores the difficulty of controlling auditory range that exists for almost all virtual auditory display systems. This difficulty stems from the fundamental ambiguity inherent in level-based range control.

A psychophysical experiment was conducted in order to determine how best to adjust the multiple relevant PSFC parameters that affect the auditory range of a virtual sound source. The ultimate goal was the development of a psychophysically calibrated control for the perceived source range using the PSFC. For all experimental stimuli, a constant value of 67 ms was set for the initial time gap between direct and indirect sound arrival (corresponding primarily to the size of the simulated space). This is well outside of the estimated 6 ms integration time of the auditory system within which indirect sound is perceptually fused with the direct sound in producing range judgments [44].

A set of 25 stimuli were formed by combining five values each of both PSFC channel volume and the PSFC liveness parameter. As was pointed out before, the liveness parameter sets the i/d ratio, and changes in the PSFC channel volume do not modify this ratio. Also recall that as liveness is reduced, the rate at which discrete reflection gain falls over time is increased, so that the reverberation time stays fixed, and the level of the last discrete reflection is made to match the level of the late diffuse reverberant field. These sophisticated controls create complicated interdependencies between simulation parameters, which presents a challenge regarding the control of virtual source range (for which no primitive parameter was originally provided). In effect, accurate control of virtual source range is confounded by variations in both the liveness parameter and in overall channel volume. Therefore, the direct empirical approach employed here was to derive a look-up table for source range based upon the results of psychophysical experiments.

3.2. Stimuli, Subjects, and Procedure

Four short speech samples, each a phonetically-rich sentence, were anechoically recorded to serve as the sound sources for range rating sessions. Table 1 shows these four sentences, and the manner in which they were presented via the two channels of the PSFC. The first and third sentence stimuli together served as a standard reference of extreme auditory range, against which the auditory range of the second and fourth sentence stimuli was to be compared. The parameters of these two comparison stimulus sentences varied from trial to trial, while the standard stimulus sentences were always presented at a fixed PSFC channel volume and liveness.

Measurements of PSFC Sound Pressure Level (SPL) were made using an integrating sound level meter (the NL-04 from Rion Co., Ltd.). Ambient background noise levels were high due to the presence of computer systems, amplifier cooling fans, and other equipment. The stimuli ranged from 46 – 70 dBA, which was confirmed by independent measurements of all four stimulus speech segments, at the five volumes employed in the second experiment (6 dB steps), and for the five liveness values employed. Due to the non-stationary nature of these speech stimuli, the physical calibration of PSFC volume control was checked using a pink noise source, and the pink noise levels set by the PSFC correlated highly with the levels measured for the speech stimuli ($r = 0.996$).

[5] German translations of terms describing perceptual attributes of spatial sound reproduction are included here as a bridge between the German and English language literature on this topic, in the belief that improved cross fertilization between these bodies of literature might benefit both.

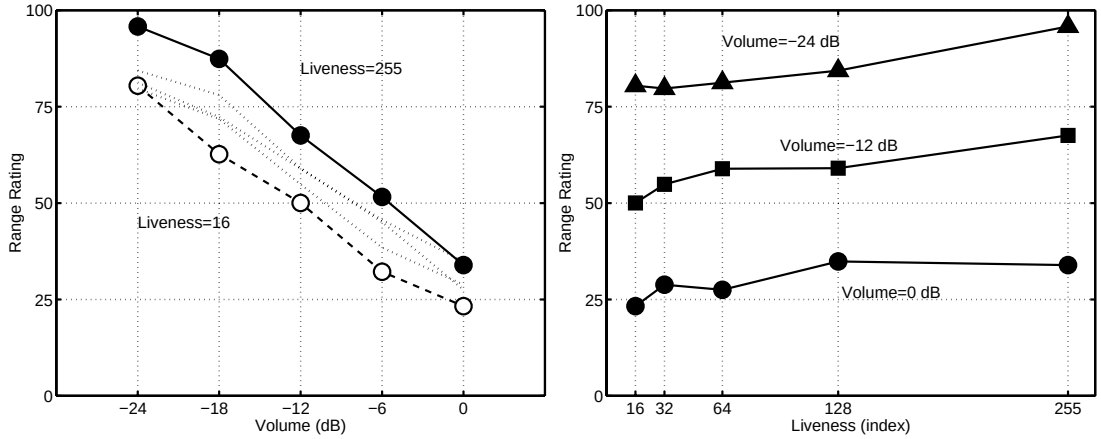

Figure 4: **Left:** Grand mean range ratings plotted over PSFC channel volume for all 11 listeners at liveness values of 16, 32, 64, 128, and 255, where the upper and lower lines represent the means associated with extreme liveness values (as labeled), and where unlabeled stippled lines show means for the intervening liveness values. **Right:** Grand mean range ratings of all listeners plotted over PSFC liveness index at volume index values of 64, 128, and 255, corresponding to -24, -12, and 0 dB.

A group of 11 listeners participated in this single listening session experiment within a one-hour period, completing all ratings twice, shifting to a slightly different position in the room between sessions. Within this single, repeated listening session, the standard stimulus was presented at a fixed volume and liveness for all trials. The channel volume index for the standard stimulus was always set to 16, corresponding to a level 24 dB below the loudest comparison stimulus, and the liveness for the standard stimulus was set to the maximum value of 255. The channel volume index values for the comparison stimuli were 16, 32, 64, 128, and 255, covering a 24 dB range, and the liveness index values were also 16, 32, 64, 128, and 255. The number of judgments in each session was 25, with a new randomized order of volume and liveness for each. Unlike in the above-described close-range experiment, the standard stimulus was not presented from the same spatial direction as the comparison stimulus, but rather from the opposite side of the room. One reason for this was to avoid a situation in which the indirect sound pattern for the two stimuli would be identical in both temporal and spatial distribution (switching sides made the simulated reflection patterns slightly different, since the listening position in the modeled space was slightly off center). Also, if the standard stimulus had been presented from the same spatial direction as the comparison stimulus, especially for such faraway sources (the range of the standard stimulus was quite extreme), some compression in the spread of responses could occur. This could happen for two reasons. First, it has been observed that a virtual sound source may reach no greater auditory range than that described by an auditory horizon when the direct sound SPL is low [45]. But also, in the absence of strong cues to auditory range, two stimuli will tend to be perceived at nearly the same range, especially if they are located in roughly the same spatial direction (Gogel [46] termed this the *equidistance tendency*).

3.3. Results

Fig. 4 summarizes the results of this psychophysical experiment, in which the comparison stimuli were presented in all combinations of five liveness values and five volume levels. The left panel shows the result of two replications for 11 subjects. Note that all range ratings here are on a common subjective scale, since they are all referenced to the same standard stimulus of fixed volume and liveness. The solid line with filled circles shows the group mean ratings at a liveness value of 255, while the thick stippled line with open circles shows the result at a liveness value of 16. The results for the other liveness values (32, 64, and 128) are shown by thin stippled lines. Again the effect of volume on range ratings is obvious, but the effect of liveness variation is more difficult to recognize. Therefore, an alternative view of the same data is shown in the graph on the right.

On the right side of Fig. 4, the group mean range ratings are plotted over the liveness values of 16, 32, 64, 128, and 255 with volume as the parameter of the graph and the group mean range ratings of all listeners are plotted over PSFC liveness. Results for only three of the five volume index values (64, 128, and 255) are shown, which extend over a 24 dB range. The upper line with triangle symbols shows that the range ratings at the lowest volume levels increase with increasing liveness. The middle line with square symbols shows the same increase at moderate volume, and the lower line with circle symbols also has nearly the same shape. The conclusion, that regardless of volume, increasing liveness causes the virtual source to move to greater auditory range, was confirmed via ANOVA: The main effect of volume was significant at $p < .01$: $F(4, 40) = 136.644$, the main effect of liveness was also significant at $p < .01$: $F(4, 40) = 10.800$, but the interaction between liveness and volume was not found to be significant: $F(16, 160) = 0.252$. In effect, the effect of volume on auditory range did not depend upon liveness.

3.4. Multiple Regression Analysis

The motivation for this psychophysical experiment was to find the relation between auditory range and PSFC parameters volume and liveness. In order to generate a look-up table for the volume level required to produce a given range rating at a given liveness value,

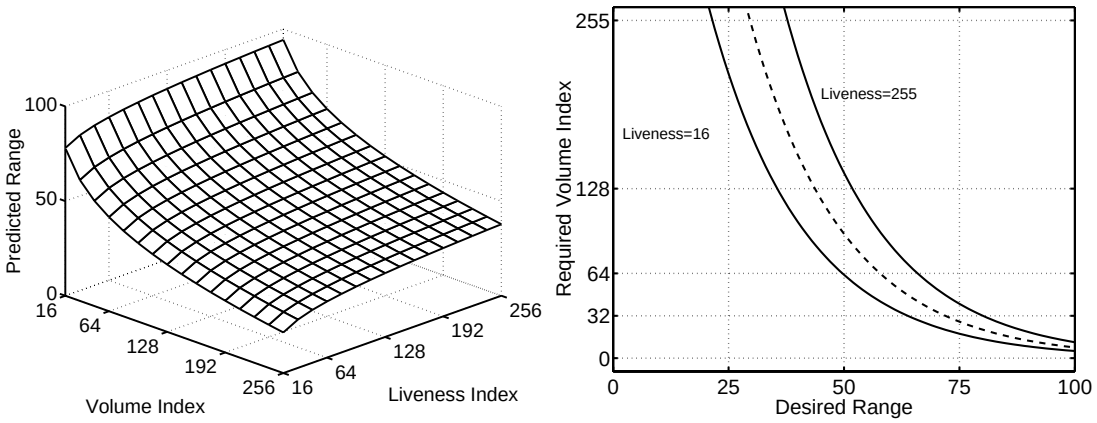

Figure 5: **Left:** Predicted Range as a function of PSFC volume index and liveness index. **Right:** Required Volume Index plotted over Desired Range for three liveness values. The two labeled curves correspond to liveness values of 16 and 255 and the unlabeled stippled curve corresponds to a liveness value of 128.

	Coefficients		95% Conf. Int.	
Model	B	St. Err.	Lower	Upper
Constant	118.127	4.258	109.763	126.491
Volume	-46.926	1.645	-50.157	-43.696
Liveness	13.532	1.645	10.301	16.763

Table 2: Coefficients (vector B) of the prediction equation fit by Multiple Regression Analysis (MRA), along with their standard error. The 95% confidence intervals for each coefficient (upper and lower bounds) are also provided.

a prediction equation was first needed for predicting the range ratings. The 550 range ratings obtained from 11 listeners in this experiment were submitted to Multiple Regression Analysis (MRA). The two predictor variables, volume and liveness, were log transformed before fitting a linear equation, so as to allow the software to fit the compressive functions observed. An alternative run of the MRA operated instead upon the log-transformed rating data, but this did not yield as good a result. As no dependence of volume effects on liveness level was observed, a simple two-variable prediction equation was fit to the data, with $R = .786$ and adjusted $R^2 = .616$. The coefficients of the "log-linearized" function are given in Table 2.

For the range of volume values and liveness values employed in the experiment, a 3D mesh plot was constructed to show how the MRA result interpolates between collected range data. The left graph in Fig. 5 shows this mesh plot with predicted range on the z axis. The desired look-up table values are shown in the right plot of Fig. 5, which values are derived by inverting the predicted auditory range function. If the user sets the PSFC liveness index to 16 and desires a range corresponding to the position 50 percent of the way to the maximum attainable range, the lower curve shows that a volume index of 64 would be required. If, however, the PSFC user sets the liveness index to 255, a volume index of 144 would be required to produce the same predicted auditory range. A different situation occurs if the user desires to obtain a range of 25 percent when the liveness is set to 255. Here, a warning message must be issued to indicate that a range of 25 is not attainable at that liveness (it is off the top of the graph!). Of course, if the liveness were set to 16, then a range of 25 percent would find a legal look-up volume value of 207. Finally, consider what the smallest attainable auditory range would be were the liveness set to its maximum of 255. Setting the volume to its maximum of 255 will bring the predicted auditory range to the 38 percent point, and no closer source is possible for a simulated space with such liveness.

4. CONCLUSION

Though auditory range can be predicted by changes in reproduced sound source level (i.e., SPL of the *proximal* stimulus), the control of source range when other auditory range cues are varied requires additional knowledge of the performance of a particular spatial auditory display system. In the case of the headphone-based display of nearby sources, a simplified model of range-dependence for simulated HRTFs needed calibration because it was based upon a somewhat unrealistic manipulation of ipsilateral and contralateral SPL. In the case of "room-based" PSFC applications, no adequate model for predicting the interaction between the channel volume and liveness parameters was available, and so an empirical calibration approach was taken. In particular, that experiment addressed the problem of adjusting for the influence of reverberant liveness and overall channel volume on judgments of auditory range. This paper taught a psychophysical calibration method employing human perceptual judgments that are first analyzed to provide a unified set of psychophysical scales for range that are reliable under known stimulus conditions. This prediction model is then inverted to produce a look-up table that can be used to set display parameter values to obtain desired auditory range percepts for the displayed virtual sources. Such look-up tables provide missing pieces of the puzzle that must be solved in the development of an API (Application Programmer's Interface) for the control of virtual source range [47]. For the two cases examined here, this information enabled the deployment of an improved API for these displays, integrated with the Sound Spatialization Framework [48] in use at the University of Aizu for a range of related spatial auditory display technologies.

5. ACKNOWLEDGMENTS

The author would like to thank Durand Begault for his review of an early draft of this paper, which helped its focus and clarified its language. Thanks are also due to Michael Cohen and Jens Herder for their comments on early drafts, and for their input throughout the course of this project. The experimental listening subjects who voluntarily participated also gave valuable feedback. This research was supported in part by the Fukushima Prefectural Foundation for the Advancement of Science and Education.

6. REFERENCES

[1] T. S. Kuhn, *The Essential Tension*, University of Chicago Press, Chicago, IL, 1979.

[2] W. L. Martens, "Uses and misuses of psychophysical methods in the evaluation of spatial sound reproduction," in *Proceedings of the 110th Convention of the Audio Engineering Society*, Amsterdam, Netherlands, May 2001.

[3] Jack M. Loomis, Chick Hebert, and Joseph G. Cicinelli, "Active localization of virtual sounds," **J. A***cous. S**oc. A**mer.*, vol. 88, no. 4, pp. 1757–1763, Oct. 1990.

[4] W. C. Gogel, "The organization of perceived space," in *Indirect Perception*, Cambridge, Massachusetts, 1997, pp. 361–386, MIT Press.

[5] P. Zahorik and F. L. Wightman, "Loudness constancy with varying sound distance," *Nature Neuroscience*, vol. 4, pp. 78–83, 2001.

[6] K. Amano, F. Matsushita, H. Yanagawa, M. Cohen, J. Herder, W. Martens, Y. Koba, and M. Tohyama., "A Virtual Reality Sound System Using Room-Related Transfer Functions Delivered Through a Multispeaker Array: the PSFC at the University of Aizu Multimedia Center," *TVRSJ:* **T***rans. of the* **V***irtual* **R***eality* **S***ociety of* **J***apan*, vol. 3, no. 1, Mar. 1998.

[7] P. Damaske, "Head-related two-channel stereophony with loudspeaker reproduction," **J. A***cous. S**oc. A**mer.*, vol. 40, no. 4, pp. 1109–1115, 1971.

[8] D. H. Cooper and J. L. Bauck, "Prospects for transaural recording," **J. A***udio* **E***ng. S**oc.*, vol. 37, pp. 29–40, 1989.

[9] D. S. Brungart and W. M. Rabinowitz, "Auditory localization of nearby sources: Head-related transfer functions," **J. A***cous. S**oc. A**mer.*, vol. 106, no. 3, pp. 1465–1479, 1999.

[10] W. L. Martens and A. Yoshida, "Psychoacoustically-based control of auditory range: Display of virtual sound sources in the listener's personal space," in *International Conference on Information Society in the 21st Century: Emerging Technologies and New Challenges (IS2000)*, Nov. 2000, pp. 288–294.

[11] D. R. Begault, *Control of Auditory Distance*, Ph.D. thesis, University of San Diego, San Diego, CA, 1987.

[12] M. Morimoto, H. Fujimori, and Z. Maekawa, "Discrimination between auditory source width and envelopment," **J. A***cous. S**oc. J**apan*, vol. 46, no. 6, pp. 449–457, 1990, (in Japanese).

[13] M. Morimoto and Z. Maekawa, "Auditory spaciousness and envelopment," in *Proceedings of the 13th International Congress on Acoustics*, Belgrade, Yugoslavia, 1989, pp. 215–218.

[14] K. Honno, W. L. Martens, and M. Cohen, "Psychophysically-derived control of source range for the Pioneer Sound Field Controller," in *Proceedings of the 110th Convention of the Audio Engineering Society*, Amsterdam, Netherlands, May 2001.

[15] P. D. Coleman, "Failure to localize the source distance of an unfamiliar sound," **J. A***cous. S**oc. A**mer.*, vol. 34, no. 3, pp. 345–346, Mar. 1962.

[16] P. D. Coleman, "An analysis of cues to auditory depth perception in free space," *Psychological Bulletin*, vol. 60, pp. 302–315, 1963.

[17] B. F. Lounsbury and R. A. Butler, "Estimation of distances of recorded sounds presented through headphones," *Scand. Aud.*, vol. 8, pp. 145–149, 1979.

[18] R. O. Duda and W. L. Martens, "Range dependence of the response of an ideal rigid sphere," **J. A***cous. S**oc. A**mer.*, vol. 105, no. 5, pp. 3048–3058, 1998.

[19] D. S. Brungart, "Auditory localization of nearby sources III: Stimulus effects," **J. A***cous. S**oc. A**mer.*, vol. 106, no. 8, pp. 3589–3602, 1999.

[20] B. B. Bauer, "Stereophonic earphones and binaural loudspeakers," **J. A***udio* **E***ng. S**oc.*, vol. 9, pp. 148–151, 1961.

[21] E. Meyer, W. Burgtorf, and P. Damaske, "Eine Apparatur zur elektroakustischen Nachbildung von Schallfeldern. Subjektive Hörwirkungen beim Übergang Kohärenz [An apparatus for electroacoustical simulation of sound fields. Subjective auditory effects at the transition between coherence and incoherence," *Acustica*, vol. 15, pp. 339–344, 1965, (in German).

[22] F. R. Moore, "A general model for spatial processing of sounds," *Computer Music Journal*, vol. 7, no. 3, pp. 6–15, 1983.

[23] R. Sommer, *Personal Space – The Behavioral Basis of Design*, Englewood Cliffs, N. J.: Prentice-Hall Inc., 1969.

[24] A. Yoshida and W. L. Martens, "Whisper function: An audio transformation for conveying a confided speech message in a multi-user virtual environment," in *IEICE Tohoku Branch Conference*, University of Aizu, Aug. 2000.

[25] N. I. Durlach, A. Rigpulos, X. D. Pang, W. S. Woods, A. Kulkarni, H. S. Colburn, and E. M. Wenzel, "On the externalization of auditory images," *Presence*, vol. 1, no. 2, pp. 251–257, Spring 1992, ISSN 1054–7460.

[26] W. M. Hartmann and A. T. Wittenberg, "On the externalization of sound images," **J. A***cous. S**oc. A**mer.*, vol. 99, pp. 3678–3688, 1996.

[27] D. R. Begault, E. M. Wenzel, A. S. Lee, and M. R. Anderson, "Direct comparison of the impact of head tracking, reverberation, and individualized head-related transfer functions on the spatial perception of a virtual speech source," in *Proceedings of the Audio Engineering Society 108th Int. Conv.*, 2000, Preprint 5134.

[28] F. E. Toole, "In-head localization of acoustic images," **J. A***cous. S**oc. A**mer.*, vol. 48, pp. 943–949, 1969.

[29] N. Sakamoto, T. Gotoh, and Y. Kimura, "On "out-of-head localization" in headphone listening," **J. A***udio* **E***ng.* **S***oc.*, vol. 24, pp. 710–715, 1976.

[30] R. V. L. Hartley and T. C. Fry, "The binaural localization of pure tones," *Physics Review*, vol. 18, pp. 431–442, 1922.

[31] D. S. Brungart, "Near-field auditory localization," in *Proceedings of the 3rd Int. Conf. on Auditory Display*, Palo Alto, CA, 1996.

[32] D. S. Brungart, "A speech-based auditory distance display," in *Proceedings of the Audio Engineering Society 109th Int. Conv.*, Los Angeles, 2000.

[33] P. A. Keating and M. K. Huffman, "Vowel variation in Japanese," *Phonetica*, vol. 41, pp. 311–322, 1984.

[34] W. C. Gogel and J. D. Tietz, "Absolute motion parallax and the specific distance tendency," *Perception and Psychopysics*, vol. 13, pp. 284–292, 1973.

[35] N. R. Draper and H. Smith, *Applied Regression Analysis*, John Wiley & Sons, New York, 1981.

[36] M. A. Gerzon, "Periphony: with-height sound reproduction," **J. A***udio* **E***ng.* **S***oc.*, vol. 21, no. 1, pp. 2–10, 1973.

[37] G. S. Kendall and W. L. Martens, "Simulating the cues of spatial hearing in natural environments," in *Proceedings of the International Computer Music Conference*, Paris, 1984, Computer Music Association, pp. 111–126.

[38] W. L. Martens and N. Zacharov, "Multidimensional perceptual unfolding of spatially processed speech I: Deriving stimulus space using `indscal`," in *Proceedings of the 109th Convention of the Audio Engineering Society*, Los Angeles, Sept. 2000, Preprint 5224.

[39] D. R. Begault, "Preferred sound intensity increase for sensation of half distance," *Perceptual and Motor Skills*, vol. 72, pp. 1019–1029, 1991.

[40] D. H. Mershon and J. N. Bowers, "Absolute and relative cues for the auditory perception of egocentric distance," *Perception*, vol. 8, pp. 311–322, 1979.

[41] J. P. Maxfield and W. J. Albersheim, "An acoustic constant of enclosed spaces with their apparent liveness," **J. A***cous.* **S***oc.* **A***mer.*, vol. 19, pp. 71–79, 1947.

[42] W. Reichardt and W. Schmidt, "Die hörbaren Stufen des Raumeindrucks bei Musik [the audible steps of spatial impression in music]," *Acustica*, vol. 17, pp. 175–179, 1966, (in German).

[43] R. A. Rasch and R. Plomp, "The listener and the acoustic environment," in *The Psychology of Music*, D. Deutsch, Ed., pp. 135–147. Academic Press, 1984, ISBN 0-12-213560-1 or 0-12-213562-8.

[44] A. Bronkhorst and T. Houtgast, "Auditory distance perception in rooms," *Nature*, vol. 397, pp. 517–520, Feb. 1999.

[45] C. W. Sheeline, "An investigation of the effects of direct and reverberant signal interaction on auditory distance perception," Ph.D. Dissertation, Department Hearing and Speech Sciences, Stanford University, 1984.

[46] W. C. Gogel, "Equidistance tendency and its consequences," *Perception and Psychopysics*, vol. 64, pp. 153–163, 1965.

[47] K. Honno, K. Suzuki, and J. Herder, "Distance and room effects control for the PSFC, an auditory display using a loudspeaker array," *3D Forum: The Journal of Three Dimensional Images*, vol. 14, no. 4, pp. 146–151, Dec. 2000.

[48] J. Herder, "Sound spatialization framework: An audio toolkit for virtual environments," *Journal of the 3D-Forum Society, Japan*, vol. 12, no. 9, pp. 17–22, September 1998.