A System to Simulate and Reproduce Audio-Visual Environments for Spatial Hearing Research

Bernhard U. Seeber[a,b,1], **Stefan Kerber**[b], and **Ervin R. Hafter**[a]

Bernhard U. Seeber: seeber@hr.mmc.uk; Stefan Kerber: stefan @hr.mrc.ac.uk; Ervin R. Hafter: hafter@berkeley.edu

[a] Audiory Perception Lab, Department of Psychology, University of California at Berkeley, Berkeley, CA 94530-1650, USA

[b] MRC Institute of Hearing Research, University Park, Nottingham, NG7 2RD, United Kingdom

Abstract

The article reports the experience gained from two implementations of the "Simulated Open-Field Environment" (SOFE), a setup that allows sounds to be played at calibrated levels over a wide frequency range from multiple loudspeakers in an anechoic chamber. Playing sounds from loudspeakers in the free-field has the advantage that each participant listens with their own ears, and individual characteristics of the ears are captured in the sound they hear. This makes an easy and accurate comparison between various listeners with and without hearing devices possible. The SOFE uses custom calibration software to assure individual equalization of each loudspeaker. Room simulation software creates the spatio-temporal reflection pattern of sound sources in rooms which is played via the SOFE loudspeakers. The sound playback system is complemented by a video projection facility which can be used to collect or give feedback or to study auditory-visual interaction. The article discusses acoustical and technical requirements for accurate sound playback against the specific needs in hearing research. An introduction to software concepts is given which allow easy, high-level control of the setup and thus fast experimental development, turning the SOFE into a "Swiss army knife" tool for auditory, spatial hearing and audio-visual research.

Keywords

binaural hearing; virtual acoustics; auralization; audio-visual interaction; hearing impairment; room simulation

I Introduction

Creating sound in a controlled way is a prerequisite for research on hearing. Headphones provide a convenient means to play sounds of all relevant frequencies and levels with high fidelity, but reproduction of the spatial information of sound fields is more problematic because individual characteristics in the head-related transfer functions (HRTFs) are difficult to capture and reproduce. Playing sound from sources in the free-field instead allows creating a predictable sound field at the subject's head position and individual ear

[1]Corresponding Author: phone: +44 115 951 8508 ext. 205, Fax: +44 115 951 8503.

characteristics are automatically accounted for. This is particularly important for work in audiology where comparison must be made between participants listening with their own ears and those listening with hearing devices where the microphone can be placed anywhere from in the ear canal to behind the ear to on the skull. Headphone presentation would not result in predictable sound pressure levels in most of those cases and free-field sound presentation is the only simple means for achieving it.

Particularly problematic for hearing impaired patients is listening in situations with multiple speakers and other sound sources present, in background noise and reverberation. Recent advances in digital technology mean that novel ways of signal processing for addressing these problems in hearing devices are possible, stimulating research into these questions. With it follows the requirement to accurately create the spatial sound field of multiple sources and reverberation at the subject's head position. The present paper shares our experience in creating loudspeaker-based setups for spatial hearing research within the context of audiology, discusses and gives guidelines for core requirements, technical and acoustical limitations and recommendations for equipment. We also discuss methodological aspects for spatial hearing research and details of best practice to implement those methods. To the best of our knowledge, there is no comprehensive overview that can guide the researcher in developing and setting up such a system for the needs of specific research projects, especially not one that highlights potential pitfalls and requirements.

Free-field presentation of sounds from loudspeakers has a long tradition in audiology where it has been used for speech testing as well as the fitting of hearing devices. However, synthesis of natural sound fields with multiple sources and reverberation via a large number of loudspeakers is rare. Implementations date back at least to 1954 when Knowles simulated selected reflections of a concert hall in an outside auditorium to recreate the hall impression (Knowles, 1954). In 1964 Meyer created an impressive setup to study room acoustics (cited in Blauert, 1997, pg. 283). Later, computers made digital simulation of reflections possible and the first free-field setup employing this technology was, to our knowledge, created by Kirzenstein (1984). In recent years, numerous room simulations have appeared in several commercial products (e.g. Odeon, Odeon A/S, Denmark) available for the acoustic design of auditoria; however, such sounds are usually rendered for headphone presentation. Yet, auralization, i.e. the simulation and reproduction of the complex sound fields of rooms, has not found its way into spatial hearing research in audiology (Kleiner et al., 1993; Vorländer, 2008). The present manuscript is a report on such systems where the focus is less on real-time simulation than on reproduction accuracy, the implementation of psychophysical methods and on easy, flexible and reliable interactions with the technology.

The Simulated Open Field Environment (SOFE) uses multiple loudspeakers to create sound sources from many different directions in the free-field as well as simulations of reverberation ranging from single echoic surfaces to complex room reflections. In the past years we have created three setups for research on spatial hearing particularly in the horizontal plane. The first was built in the anechoic chamber at the Technical University in Munich, with eleven custom loudspeakers and the possibility for the listener to interactively position a visual marker to display the perceived direction of the sound (Seeber, 2002). It was used to study the localization ability of patients with cochlear implants (Seeber *et al.*, 2004; Seeber and Fastl, 2008) and to refine virtual acoustics (Seeber and Fastl, 2003). The second was in the anechoic chamber of the Psychology Department at UC Berkeley where 48 loudspeakers allow simulation of reverberant sound fields of rooms while visual stimuli projected on a screen in front of the participant can be used to simulate a visible space and/ or support interactive manipulations by the subject. It was used to study the effects of reverberation with different patient groups (Seeber et al., 2008; Seeber and Hafter, 2007b) and the role of HRTFs in interference patterns due to acoustic panning (Seeber and Hafter,

2007a). It followed a previous setup which used a smaller number of loudspeakers to create a sound source and a single reflection to investigate, e.g. the lateralization efficacy of portions of ongoing stimuli (Stecker and Hafter, 2002). The third setup is a refinement and extension of the second at the MRC Institute of Hearing Research in Nottingham. It uses custom loudspeakers to increase the dynamic range and accuracy of sound reproduction and multiple projection screens to display visual spatial stimuli and video sequences, a facility now used for the work with children.

In this paper we focus on ways to achieve accurate sound reproduction as needed for research including: the need to play sounds in an anechoic chamber, the effect of loudspeaker size, distance from speakers to participants and electronic equalization. We also present the design of a specific loudspeaker for the SOFE and software concepts that allow for easy development of experiments. Unique to research in spatial hearing we will discuss useful methods for collection of localization responses and their implementation in an interactive visual environment. We will conclude with an overview of spatial hearing experiments that make good use of this setup.

II System description

II. 1 Overview

A system to interactively generate and play audio-visual stimuli in space requires a custom approach which is introduced in this section while its individual components are discussed in the following sections. Students and other temporary staff are often involved in research which requires that experiments be generated flexibly, easily and in short time after only brief training. At the same time, an auditory-visual experimental system is, by its nature, complex. The 48-channel audio system we discuss here presents a high data rate of 135 MBit/s out of Matlab and equalizes loudspeakers on-the-fly, while using a visual display system that is synchronized to the audio as well as providing interactive, real-time feedback on multiple video screens. And, of course, work with patients demands that the system runs reliably.

Figure 1 shows an overview of the hardware, the communication links between devices and the signal chains. The system follows a modular approach, with three computers handling the audio part, the visual part and input devices, respectively. All sound playback, recording and equalization are handled on one personal computer (PC) which also exerts central control of the overall system. This allows listening tests without interactive input or visual feedback to run on a single computer, making programming simple for a wide range of experiments. Sound playback and system control are handled by custom functions in Matlab (Natick, Massachusetts, USA), a programming language which makes matrix-based processing of multi-channel audio simple.

The visual environment is self-contained and installed on a separate PC with a multiple-output graphics card. It is controlled by data messages on the computer network, e.g. from the audio PC, to display certain objects or to position them at calibrated azimuthal and vertical positions. The message exchange is fast compared to the visual frame rate and thus achieves synchronization between the audio and video PC.

A third computer handles input devices, e.g. a trackball, and sends messages to the audio and video PCs for interactive changes in stimuli. Handling stimulus inputs separately from the audio and the visual environment is useful because it simplifies modifications to either of them. Because specific experiments only need certain system functionalities, the modular approach allows those which are not needed to be turned off, resulting in a higher stability of the overall system.

II.2 Sound reproduction

Loudspeaker Requirements—The reasons for using a loudspeaker for spatial hearing research were laid out in the introduction. The core requirements are to reproduce sounds from various directions with high temporal accuracy at defined, high sound pressure level across the frequency spectrum of interest. Physical and practical limitations pose restrictions on those requirements and this section will highlight design criteria for the loudspeaker setup.

The required output power of loudspeakers and amplifiers depends mainly on the following aspects:

1. The sound pressure level required at the subject's head: For most psychoacoustical research with normal hearing subjects, a level of 70–80 dB SPL at the head seems sufficient. However, more may be necessary to ensure audibility with hearing impaired patients. For mild-to-moderate hearing impairment, 90 dB SPL may be sufficient, while 110 dB SPL may be required for comfortable loudness with more severe hearing impairment.

2. The distance d of the loudspeakers from the listener: Loudspeakers should be placed such that the listener sits in the far field, where frequency response, levels and interaural level differences (ILDs) are predictable within a range around the listener position. This is generally achieved when the distance d is larger than the wavelength λ of the lowest frequency f reproduced, $\lambda = c/f$, $c \approx 344$ m/s, which for the average fundamental frequency of male voice of 100 Hz results in d > 3.4 m (Kuttruff, 2007). For practical purposes a distance of half a wavelength may be sufficient, reducing the requirements to a minimum of 1.7 m. This is in line with Krebber at al. (1998) who evaluated the results of the AUDIS project and recommended 1–3 m distance to measure HRTFs and the limit of 1 m refers to a cut-off at 200 Hz. Brungart and Rabinowitz (1999) investigated HRTFs for distances below 1 m and showed that the violation of the far field leads to substantial ILDs which can serve as – here unwanted – distance or localization cues.

   It should be further considered that the 1/d reduction in level leads to better level predictability at larger distances. Suppose that accuracy of the participant's head position varies during a session, say, for example by $\Delta d = 10$ cm. If so, the change in level from 1/d to 1/(d+10 cm) is much less for larger distances d. According to $\Delta L = 20 \log_{10} (d+\Delta d/d)$, a minimum distance of 0.82 m is needed to achieve an accuracy better 1 dB. However, not only test room size might limit loudspeaker distance. Changing distance from 1 m to 2 m requires sounds be played 6 dB higher for the same SPL at the position of the listener.

   In summary, loudspeaker distance d should be larger than 0.5 times the largest wavelength, i.e. d > 0.5 c/f. If this condition cannot be met, sound fields cannot be reproduced accurately at the lower frequency and high-pass filtering of the stimulus should be incorporated.

3. Provision of headroom for loudspeaker equalization: Frequency equalization to compensate the frequency response of the loudspeaker can require substantial amplification at selected frequencies. Small loudspeakers generally act as bandpass filters, doing reasonably well with mid-frequencies, but attenuating low and high frequencies. However, it is often important in psychophysical experiments that the spectral weighting of sound components be consistent, so that masking remains predictable (Zwicker and Fastl, 1999). This requires that the frequency response of the overall system, especially the loudspeakers, be independent of frequency, i.e.

"flat". That can be achieved by filtering the sound with the inverse loudspeaker or system impulse response prior to playback as discussed below. In accord with our experience, the amplification to equalize the frequency response should not exceed 6–10 dB. Given this, one should note that 10 dB additional amplification requires a factor of ten higher output power from the amplifier. Suppose a loudspeaker delivers 90 dB SPL per Watt at 1 m distance. If 10 dB amplification is needed to maintain 90 dB SPL across the spectrum the amplifier must supply 10 Watts.

4. Provision of headroom for fluctuating and impulsive sounds: The narrow band signals and impulsive, click-like sounds often used in the study of the precedence effect have a high crest factor, i.e. their maximum level far exceeds the long term average. For the loudspeakers to reproduce the instantaneous maximum level without noticeable distortion requires 20–30 dB of headroom. Since those sounds also evoke a lower loudness compared to broadband signals of longer duration, additional amplification may be needed to compensate, further exacerbating the requirement for headroom (Zwicker and Fastl, 1999).

Designing a loudspeaker setup given these restrictions means finding an optimal compromise between counteracting requirements. Suppose that we make a worst-case estimate for the level requirements to achieve 90 dB SPL at the subject's head:

Nominal sensitivity:	90 dB SPL per Watt at 1 m distance
Distance:	+ 6 dB if loudspeaker is placed 2 m from the subject
Headroom (equalization): + 10 dB	
Continuous level at 1 m: 106 dB SPL	

Headroom (crest factor): + 20 dB	
Peak level at 1 m: 126 dB SPL	

According to this estimate, the maximum level the system should be able to reproduce without pronounced distortion is 126 dB SPL in 1 m distance. The extra 36 dB above the nominal loudspeaker sensitivity requires the amplifier to deliver 3980 W power and the loudspeaker to handle it – practically impossible! However, signals with high crest factors are broadband so that the amplification of low and high frequencies due to speaker equalization does not lead to the same requirements in signal amplitude as found for sinusoids. Thus, for most sounds, reducing the total headroom to 15–20 dB will be sufficient. Nevertheless, including distance weighting this requires amplifiers and loudspeakers to handle a power of 100–400 W. It is thus very important to assess level requirements critically.

Another important design criterion for the loudspeaker setup is the frequency range it covers. The loudspeaker frequency range should not be considered separate from other factors influencing the sound-field at the listeners head, for example those relating to the test room. Sound absorption in the room is reduced at low frequencies, leading to standing waves and thus unpredictable level and binaural cues at the subject's ear. Only large, high-quality anechoic chambers achieve anechoic conditions, that is a level predictability of ± 1 dB, down to 100 Hz (EN ISO 3745: 2003). The level variability in most audiometric test booths will thus be sufficiently larger and care should be taken to assess how variability is tolerable in a given experiment. If the experiment would be sensitive to small changes in level at low frequencies as, for example, those related to recruitment (Moore, 2007), the upward spread of masking or accurate presentation of binaural cues, then it might be better to high-pass filter stimuli appropriately to ensure predictable conditions. We have already discussed the need for a sufficient distance to the loudspeaker to achieve predictable far-

field conditions and if they cannot be met, stimuli should be high-pass filtered. With this said, there is no need to incorporate loudspeakers made to produce the lower frequencies. A third consideration is that the frequency range of the loudspeakers does not need to exceed the bandwidth of the test sounds or, in fact, the hearing range of the listeners. Speech, for example, does not contain pronounced energy at the fundamental frequency which is as low as 100 Hz. On the other hand, one may want to play low frequency noise to mask difference tones in masking and binaural detection studies (Wegel and Lane, 1924), while high frequencies should be reproduced when studying elevation perception. In summary, loudspeaker bandwidth requirements should be carefully assessed against the experiments planned, stimuli used and the acoustical limitations of the test room.

Loudspeakers should be as small as possible to achieve high sound field accuracy. However, this conflicts with the requirement for a high output level to test patients without their hearing devices because high output levels can only be achieved with large loudspeakers. Unfortunately, large speakers cannot be closely spaced and have difficulties in reproducing high frequencies where hearing loss for most hearing-impaired people is predominant and a compromise between power and high frequency capabilities must be found. Additionally, the surface of every loudspeaker reflects sound from other loudspeakers in the room, thus making an anechoic environment more echoic. Larger loudspeakers reflect more energy, particularly towards lower frequencies, both reasons to design them as small as possible. These reflections might be more pronounced in circular arrangements as loudspeakers are opposite each other and directed to the listener who sits close to the focus point of all reflections. Loudspeakers with high sensitivity should be considered as they deliver higher acoustic power from a certain electric input, potentially making the choice of a smaller loudspeaker possible.

It is a difficult task to construct a single loudspeaker which covers a large frequency range and delivers high output power. To overcome this problem, commercially available loudspeaker systems often consist of multiple loudspeaker chassis, each covering a certain frequency range. This can be problematic for spatial hearing research: The different speakers are located at different positions inside a common housing; thus different frequencies originate from different positions – a drawback for localization studies unless loudspeakers are placed very far from the subject. Similarly, many speaker systems rely on the bass-reflex principle, where openings are brought into the housing to enhance the bass response. Those openings dissipate sound from other than the loudspeaker membrane location. The different speakers in a combined housing are connected using a filter-network which assures that each is supplied with a particular frequency range. Although this allows for a flat frequency response over a wide frequency range, two loudspeakers are engaged at the cross-over frequency, leading to difficult to predict phase transitions. This will normally not be critical in studies of speech-in-noise or for localization of isolated sounds, i.e. "standard" audiological tests, but it may be critical in studies where temporal information within or across auditory filters is evaluated, such as those on auditory grouping, co-modulation masking release, and binaural masking level differences. Another critical application is when a sound field is constructed by playing from multiple loudspeakers, where differences in each loudspeaker's level and phase response can lead to substantial alterations of the summed sound at the subject's ear. Examples are precedence effect scenarios or panning between loudspeakers (Seeber and Hafter, 2007a).

A Customized Loudspeaker Design for Spatial Hearing Research—Most commercial loudspeakers are two- or three-way systems based on the bass-reflex principle and build in a rectangular box. This leads to several drawbacks for their use in spatial hearing research as outlined above. The loudspeakers chosen for the setup are made of a single loudspeaker chassis mounted in a small, closed, round housing. This minimizes sound

dissipation to and reflections from neighboring loudspeakers. Phase effects based on summation of the sound from the loudspeaker chassis and from openings in the housing are avoided, thus resulting in a better approximation of a point source. The loudspeaker box Peerless, model 821385 (Tymphany Corporation, Denmark), combines these criteria and was chosen for the setup in Berkeley (Hafter and Seeber, 2004). However, it exhibits a strong mid-frequency bandpass-characteristic, making significant amplification of low and high frequencies necessary which limits the maximum possible sound pressure level.

For the setup in Nottingham a custom loudspeaker was developed that is better suited for the work with hearing impaired participants. Loudspeaker chassis with a diaphragm size of 5–6 inches can fulfill the level and frequency requirements and thus provide a good compromise for the SOFE application. We decided on a H1262 midrange loudspeaker chassis made by SEAS (Ryggeveien, Norway) because it can function in a small, closed housing, it has a sensitivity of 90 dB SPL per Watt, it handles a maximum power of 300 W and it shows a relatively wide, flat frequency response. Following the recommendations by Small, a loudspeaker housing with an effective volume of 1.71 liters was constructed to yield a "2nd - order Butterworth maximally-flat alignment" (Small, 1972; Small, 1973). This leads to a minimum overshoot in the step-response of the speaker whilst keeping the frequency response as wide as possible. A sketch of the constructed housing is shown in Figure 2.

The construction uses standard parts wherever possible for ease of construction and cost effectiveness. A round construction was chosen in order to diffract sound-waves reaching the outside of the housing. The outside shell was made of 160 mm polyvinyl chloride (PVC) pipe (sewage pipe), closed by medium-density fiber board at either end. Front and back are airtight sealed and the front board holds the loudspeaker chassis. To suppress standing waves inside the housing the back was cut at an angle of 80 degrees and the inside was filled with polyester wool.

The frequency response of the loudspeaker construction is flat within ±6 dB from 170 Hz to 10000 Hz as shown in Figure 3 (dotted line). This frequency range is sufficient to approach most of the common research questions in spatial hearing. Additionally, it shows only moderately steep transitions between minima and maxima, important for equalization with short digital filters as described below.

Loudspeaker setup: Mounting and electric connection—The setup in Nottingham consists of 48 loudspeakers, 36 of which are mounted horizontally at ear-height (124 cm) with an azimuthal separation of 10°(Figure 4). Horizontal speakers are mounted on poles (2 cm iron rods) which are clamped to a railing (5 cm × 5 cm steel) running along the walls of the anechoic chamber. One speaker was mounted on a lever attached to the railing to avoid obstruction of the door whilst preserving the 10° separation. The use of a railing along the rectangular chamber walls has the following advantages: (i) it avoids complicated circular parts, (ii) it places loudspeakers not in a circle which avoids reflections from the loudspeakers summing up constructively at the subject's position, (iii) it makes it easier to put up a flat projection surface for a visual display, and (iv), the listening position can be accessed without moving any of the loudspeakers which is not possible when many speakers are placed in a tight circle, thus making it easier to guide patients to the seat. It is important to note that this mounting principle at different distances to the listening position introduces different levels and time delays. However, they are static and frequency independent and thus relatively simple to correct, e.g. by increasing the volume at the amplifier and delaying the signals, effectively making the setup a circle. The correction can be incorporated in the loudspeaker equalization routine described below and needs to be developed only once, thus it seems not too costly compared to the gain in easier everyday access. It nevertheless

appears a matter of taste if one prefers a circular arrangement for simplified math or a rectangular arrangement with the particular access benefits.

Easily detachable and movable clamps connect the speaker poles to the railing (Figure 2). Since each loudspeaker is balanced on top of a pole it forms an inverted pendulum which can move when playing sounds or from accidentally hitting the loudspeaker or the railing. This is damped by custom pads in the clamps (Aearo E-A-R Isodamp C-1002 in Berkeley and Farrat NBR 40–8 in Nottingham).

In Nottingham, additional twelve loudpseakers were installed at elevated angles of ±40° for playback of floor and ceiling reflections of simulated rooms. The elevations were calculated to yield roughly minimum spectral error in human HRTFs across the range of elevations found in our room simulations, i.e. the average spectral error for elevation cues is kept small when sources from within ±20° off the horizontal plane are mapped to horizontal speakers while sources outside this range are mapped to the nearest speaker at ±40°. These loudspeakers are either hung into the wire-mesh floor of the anechoic chamber or are attached to an iron frame at the ceiling at azimuth angles of 0°, ±50°, ±130° and 180°

Loudspeakers are connected via shielded cables (core diameter 2 mm^2) to 24 amplifiers in the control room. The front, side and elevated loudspeakers are connected to Alesis RA-300 amplifiers, delivering 90 W into 8 Ω, and the most distant rear speakers are driven by the more powerful Alesis RA-500 amplifiers (150 W into 8 Ω, Alesis, Cumberland, USA). The setup in Berkeley uses Crown D75 amplifiers (35 W into 8 Ω, Crown Inc., Elkhard, USA) which proved to be not powerful enough for some test sounds from rear speakers. The amplifiers are connected to Motu 24 I/O sound cards via symmetric cables (2 modules with 24 channels each, 24 bit, 96 kHz, MOTU Inc., Cambridge, MA, USA).

Additional electrical design may be necessary. Despite using symmetric connections and studio quality amplifiers, the system with Alesis amplifiers produced audible hum. Measurements revealed that the amplifiers' right channels were generally noisier than the left channels. The following modifications were applied to the Alesis RA-300 amplifiers: (i) The shield of one of the audio wires connecting the potentiometers to the main circuit board was cut, (ii) a piece of μ-metal was mounted between the circuit-board responsible for the right channel signals and the power transformer inside the housing, and (iii), the top cover of the amps was replaced by an aluminum cover. The modifications reduced the hum in the left and right channels by 1dB and 5 dB, respectively. Additionally, the ground wire in the symmetric input connection from the soundcard was disconnected, leading to a further reduction of hum, together rendering hum inaudible.

Loudspeaker measurement, calibration and equalization—While one must maintain control of the sound pressure levels in any psychoacoustical experiment, an additional problem in the SOFE is the need to be aware of the phase response of each loudspeaker. This can be especially important when playing sounds simultaneously from multiple loudspeakers as, for example in studies of the precedence effect. For this, the speakers should be equalized to the same, preferably linear, phase response as measured at the position to be occupied by the centre of the listener's head.

The calibration and equalization are done in three steps: First, the amplification of each amplifier is adjusted such that sounds achieve roughly the same level at the listeners head independent of speaker distance, thus giving each channel roughly the same digital headroom. Because the amplification is set to the just necessary amount which is lower for nearby speakers, the audible hiss of the system is also minimized. The calibration is done with the help of a custom-made switcher unit. The switcher is remote-controllable from the

audio-PC and allows routing the output of each amplifier channel to a voltmeter. After manual adjustment of amplification, a fast automated routine can read out the voltage for probe tones on each channel and compare it to predefined target values. This is convenient in day-to-day use to alert the experimenter if amplification changed in a particular channel.

The second calibration step achieves an accurate overall level calibration and equalizes each channel for time delay and amplitude response. A recursive algorithm generates 512-taps finite-impulse response (FIR) filters (at 44.1 kHz sampling frequency) which are embedded in 1024 taps filters, leaving the other 512 taps at zero. The effective filter is still 512 taps long and the extra zeros only create computational overhead. However, temporally shifting the 512 taps within the 1024 taps filter is an easy means for compensating time differences between loudspeakers. The filter length was confined to 512 taps because the auditory system could otherwise temporally "hear into" the changes applied over the length of the filter through its ability to conduct a time-frequency analysis.

The measurement signal is played from Matlab via soundcard, amplifiers and loudspeakers. A ½-inch measurement microphone with pre-amplifier (G.R.A.S. 40AF microphone, 26AK pre-amp, 12AK power supply and amplifier, G.R.A.S. Sound and Vibration, Holte, Denmark) is hung from the ceiling to record the sound signal at the location of the center of the participant's head. This is fed to the analog input of the soundcard. A microphone with identical, flat transfer function for all loudspeaker directions should be chosen.

The algorithm first finds a filter that equalizes the frequency response of the loudspeaker and assures identical, calibrated levels across speakers. A maximum-length sequence (MLS) longer than the expected room impulse response is used to measure the initial frequency response by cross-correlating the original and the recorded sequence (Rife and Vanderkooy, 1989). An inverse filter of 512 taps length is generated by frequency-limiting and inverting the spectrally smoothed and subsequently downsampled magnitude spectrum of the measured impulse response. This filter is then applied to Gaussian noise band-limited to the target frequency range of the system. The filtered noise is played next to measure the overall level and the amplification of the filter is adjusted through comparison to a previously stored reference level. The procedure is repeated several times, with the difference that subsequent MLS-measurements correct the filter frequency response by the inverted error from the target frequency response.

Having generated filters to assure equal levels and amplitude response from all loudspeakers, the algorithm then finds their temporal offsets to an accuracy of a few samples to equalize for different loudspeaker distances. The offsets are used to shift the 512-taps FIR filters generated above within 1024-taps filters. The second equalization step can be done *without* synchronization between playback and recording and thus in *any* system, because only the spectra and relative time differences between channels are evaluated. This equalization achieves what is needed most: calibrated, identical levels across all loudspeakers, frequency-independent response and equalization of the distance delay to an accuracy of ~100 μsec.

The equalization routine was later extended by a third step which additionally equalizes the phase response of each loudspeaker and corrects the remaining distance delay error, but relies on synchronization between playback and recording. Clicks are filtered with the previously obtained filters, played, and the phase response is determined from their recordings, inverted and applied to the filter. The temporal windowing of the FIR-filter introduces a level error which is corrected in a subsequent measurement identical to the first step. After the third step, filters are obtained which also correct the phase response, e.g. correct for the tendency of the loudspeaker to respond more sluggishly at some frequencies

as well as to correct for phase response changes around resonance frequencies, resulting in temporally sharp presentation of transients. Further, time delay errors between speakers are corrected to a few µsec. Figure 3 shows equalization of the frequency response to ±1.5 dB in 200 Hz – 12 kHz. Deviation of broadband level between loudspeakers is smaller than 0.3 dB for steady as well as impulsive sounds.

Software for sound payback—Equalization filters are incorporated in a custom sound driver for Matlab which applies an overlap-add FFT-algorithm (fftw-library, www.fftw.org) in real time during playback for all 48 channels. This has the advantage that filters obtained immediately prior to the experiment can be used without time-consuming off-line filtering of all experimental sounds. The sound driver uses the ASIO interface (Steinberg Media Technologies GmbH, Germany) which bypasses the Windows sound system and does not permit the user to change the output volume, a prerequisite to maintain calibrated levels.

II.3 Visual display

Hardware for visual projection in sound test rooms—A temporally and spatially accurate display of visual information makes it possible to study auditory-visual interactions such as the ventriloquism effect or lip-reading benefits. The option to present visual information to the participant can further be used to give visual feedback or to display instructions. Visual projection can thus be a very useful add-on to a hearing test facility. This section reviews requirements and gives advice to resolve problems that might be encountered when setting up complex audio-visual environments.

The simple case of placing a computer screen in front of the participant to gather responses or to provide feedback needs no further explanation. If the visual information is meant to integrate with the three-dimensional auditory environment of rooms, spatially accurate reproduction of visual information is needed. This can be best achieved by projecting the visual image on screens in the test room. The suitability of a projector for this application is determined by several factors: resolution, noise level, throw ratio and brightness/black level.

If visual projection should cover a wide angle of view, high resolution is needed to display text sharp and without visible pixelation. This is important for displaying instructions or when programming and piloting experiments whilst sitting inside the test-room. Current high-definition projectors have a resolution of 1920×1080 pixels which results in a pixel size of 1.1 mm on our 2.1 m wide screens. Pixelation of letters is minor and visual objects can be placed with high accuracy. For example, localization studies using a light-pointer method (e.g. *ProDePo-method,* Seeber, 2002) can therefore be done with a theoretical accuracy of 0.06° which is well above the localization accuracy of the human auditory system under optimum conditions (Blauert, 1997).

A problem with most currently available projectors is their insufficient "throw ratio": Projectors do not allow for a big enough picture given the short projection distances encountered, e.g. in an anechoic chamber. The system in Nottingham uses a special lens to magnify the image 1.5 times which allows for 0.67 times closer placement of the projector to the screen whilst preserving the image size (Navitar, Rochester, USA; Type SSW065). This also permits the projector to be placed closer to the participant to prevent the participant's head from interrupting the light path.

Two further factors characterizing projectors are "brightness" and "contrast". Tests of spatial hearing are often done in complete darkness to prevent a "visual frame of reference" (Shelton and Searle, 1980). Because of the dark adaptation of the eyes, projectors are not required to have a high brightness; however, if testing needs to be done with ambient light, e.g. for testing children or patients with vestibular problems, high brightness projectors are

needed. On the other hand, if the room has to be dark during an experiment, a projected black image has to be dim enough to not introduce unwanted light. This is difficult to achieve and projectors with high contrast are needed.

The lamp in the projector dissipates a fair amount of heat and fans are needed to cool it. The fan noise can interfere with auditory experiments and the projector should be as quiet as possible. The effective noise can be reduced by placing the projector as far as possible from the participant, which, however, may bring the head in the light path. Alternately, the directional component of the noise can be removed by mounting it directly above the subject, albeit at the cost of an increased noise level and thus auditory masking. When arranged as in Figure 4, the three Epson EMP TW-2000 projectors of the Nottingham system create a noise floor of 22 dB(A) at the subject's head position. Some additional issues have to be considered regarding the connection of devices, the graphics card and the projection screens which are described next.

Transmission of images with high resolution from the computer in the control room to the projector in the anechoic chamber is critical because a high amount of data is sent over long cables, in our case up to 15 m. Connecting the video projectors via a digital link (High Definition Multimedia Interface, HDMI) allowed for the necessary cable lengths at high data rates without the usage of repeater devices. Additionally, the digital video-signal can be easily distributed to multiple monitors using a HDMI-splitter (Lindy Electronics Limited, Stockton-on-Tees, England) which allows the experimenter to monitor in the control room exactly what is shown to a test person.

Unfortunately, long signal paths and the use of signal splitters can cause synchronization problems with the graphics card of a computer. Some cards need to recognize the projector at startup of the PC, otherwise their output cannot be switched on later. An additional device connected directly to the output of the graphics card emulates the presence of the projector in cases where detection fails (DVI Detective, Gefen, Chatsworth, USA).

The graphics card should be chosen not only for processing speed, memory size and number of outputs but also for its ability to directly interpret the commands from the visual environment in order to maximize speed. Because our environment was programmed in OpenGL, an OpenGL capable graphics card with four outputs was chosen (NVidia Quadro NVS 440; NVidia, Santa Clara, California, USA).

Projection screens inside the anechoic chamber need to be visually opaque but acoustically transparent while presenting images clear and sharp. Special fabric exists for cinemas, but was not chosen due to cost constraints and our special needs for the SOFE setup. The possibility to roll up the curtains during experiments not involving visual presentation constitutes such a need. The frame from a roller blind was used and the original fabric was replaced by white "spandex", a synthetic fiber with extraordinary elastic properties. Informal comparison tests showed this material to be the best choice in terms of acoustical as well as optical properties. The curtain should be mounted as close as possible to the loudspeakers in order to minimize any potential effects of standing waves. Loudspeaker equalization can then fully correct for the slight attenuation at high-frequencies due to the presence of the projection screens.

Software for visual rendering—The creation and display of visual content is handled in a stand-alone, independent program which is controlled via messages on the network. Keeping the visual rendering separate from the audio part is beneficial in several ways: 1) it allows full parallel processing which is not only needed due to the high processing load for playing and equalizing 48 channels of audio but also allows more complex visual scenes to

be created, 2) the visual renderer does not need to be started for audio-only experiments, 3) the modular approach is more flexible and easier to program.

The creation of visual content can be done with a multitude of libraries and programming environments, e.g. in Python using OpenGL, with one of the 3D-game engines (e.g. "Unreal engine"), with MAX/MSP and jitter, or, like in our case, with a C/C++ code which uses the OpenGL and GLUT libraries. The latter is a low level approach which allows for fast rendering of visual content at the cost of image complexity.

The visual renderer receives messages in the Open Sound Control (OSC) format (http://opensoundcontrol.org) which are interpreted immediately to ensure a change is visible in the next displayed frame. The messages for example turn objects on or off and position them, change the object texture or the background. The renderer can be switched to accept or reject messages from certain input devices making it possible to position objects with a selection of multiple input devices. For example, the *ProDePo*-localization method in which participants position a visual pointer to the location of an auditory event can be implemented as follows: After the sound has been played from the audio PC, a message is sent to the visual renderer to display a light spot in front of the subject and to allow its position to be controlled from a trackball. The subject can then turn on the trackball and the trackball software sends position change messages to the visual renderer. The visual renderer moves the light spot across the screen according to the messages it receives, i.e. according to the participant's input, until the participant presses a button. Upon receiving the button message the visual renderer registers the location of the light spot, turns it off and sends the location to the audio-PC where it is recorded as the localized position. This process has the benefit that the audio-PC has nothing else to do during the pointing process but to wait for the message with the localized position – time that can be used to process or load the next sound. Additionally, because the processes necessary for moving visual objects reside outside the relatively slow Matlab environment, responsive interaction with visual objects is possible. To create smooth movements of visual objects, an update rate of at least 30 Hz is required which exceeds Matlab's capabilities. The modular approach of the system also allows position messages to be sent from other input devices, e.g. a Polhemus position tracker.

Synchronization between audio and video can be done on a trial-by-trial basis for most auditory experiments which present single events rather than continuous streams. Since the visual renderer interprets a message within milliseconds, the visual content is changed, usually in the next frame, with frames occurring every 16.7 ms at 60 Hz frame rate. If necessary, a higher accuracy could be achieved by synchronizing the audio playback to the visual frame rate across the two computers. If high synchronization accuracy is required, it should be kept in mind that each stage, buffering of audio and video, sound equalization, travel of sound in the free field, video transmission over the serial HDMI connection and the display of the picture in the video projector, has its own delay which must be accounted for.

II.4 Room simulation

Simulating the impulse response of rooms for use in auditory tests has the advantage that room conditions can be carefully chosen for a specific test and that even parameters of individual reflections, e.g. their level, can be altered. The aim of using the room simulation in conjunction with a multi-loudspeaker setup is not only to accurately reproduce the level of reflections – this can be done over headphones – but particularly their spatial distribution for study of possible benefits of spatial hearing. What is more, the loudspeaker-based simulation remains reasonably correct for head turns and small head movements and is useful for work with patients using hearing aids or cochlear implants. Individual reflections are played from a loudspeaker close to the direction of their respective reflection-location in

the simulated room. Since the setup has a higher speaker density in the horizontal plane, binaural cues for horizontal localization can be accurately reproduced, while elevation cues are less well simulated, in line with poorer localization ability for elevations. The auralization process for the SOFE setup includes the following:

1. The location of reflections, their level, time delay and spectrum is determined with room simulation software. They depend only on the room geometry, wall materials and the location of source and receiver, but not on the sound reproduction system. Our simulation is based on the mirror-image source method and allows simulation of rooms with arbitrary shapes which achieves a good approximation of early reflections (Bolt et al., 1950). Mirror-images are back-traced to ensure their validity and visibility. The computational complexity is reduced at high reflection orders by cutting branches that lead to invalid mirror-images over several subsequent reflection generations. Mirror images are also temporally jittered to increase the naturalness of the simulation (Deane, 1994; Kirszenstein, 1984; Lee and Lee, 1988; Vorländer, 2008).

2. Individual reflections are mapped to one or more loudspeakers for sound reproduction. A separate filter is created for each loudspeaker with the loudspeaker's contribution to the overall room impulse response. Reflections are mapped either to the nearest speaker or they are panned between speakers. A specific panning technique was developed which creates no audible panning errors for nearby, horizontal speakers (Pulkki, 1997; Seeber and Hafter, 2007a). In experiments, the direction of the direct sound is chosen such that it matches the position of a loudspeaker. This has the advantage that the important direct sound is played from a single loudspeaker, thereby avoiding position dependent interference (comb filter effects) and the large level variability at high frequencies which occurs when playing from multiple loudspeakers in Ambisonics (Gerzon, 1973) or wave field synthesis (Berkhout et al., 1993).

3. The sound is convolved with the filters containing the room impulse response contributions of each loudspeaker and played.

Commercial room simulation software will excel in step 1, however, step 2 is often not possible because the reflections are rendered internally, e.g. with head-related transfer functions, and the user has no options for changing it. Implementing the mirror-image source method for research purposes gives the experimenter control over each individual reflection, a prerequisite for studying their influence.

III Application Areas

The flexible SOFE-system can be used for various experiments on spatial hearing, listening in noise and reverberation and on auditory-visual interaction. Tests of auditory localization can utilize the visual environment to gather pointing responses to sound positions. Subjects are instructed to position a visual object, e.g. a dot or a picture of a loudspeaker, to the perceived position of the sound by turning on a trackball (Seeber, 2002). An indication of perceived distance is for example possible if the visual environment shows the 2D-projection of the 3D-image of a room.

Baseline auditory localization can be tested in anechoic conditions and further tests may add single or multiple reflections of simulated rooms. One example is testing localization dominance in the precedence effect: A leading sound may be played from a loudspeaker at, e.g. $+30°$ and accompanied by its delayed copy from $-30°$. Participants are then asked to point to the perceived direction of the sound which may be, dependent on the delay, between the loudspeakers in accord with summing localization for very brief delays, in the vicinity of

the lead loudspeaker given precedence or, for still longer delays, beyond the echo threshold, as a separate pair of sounds (Litovsky *et al,* 1999; Seeber and Hafter, 2007b). More realism than playing a single reflection can be achieved by playing multiple reflections of simulated rooms. This can be used to address current questions of binaural hearing in rooms particularly for patients using hearing devices. Accurate comparison between normal hearing participants and those using hearing devices is possible because both sit in the same, free sound field generated by the loudspeakers.

The visual environment can not only be used to collect feedback from the participant, but also to provide feedback to the participant for example in experiments on auditory training where presented auditory directions could be indicated visually. The visual environment is also useful for working with children who, like in visual response audiometry, can be engaged in the task by providing appropriate feedback in the form of pictures, comic characters and video sequences.

The SOFE setup further allows for studies on auditory-visual interaction and its role in object formation, e.g., as in ventriloquism. Sound sources can be played with high spatial accuracy, particularly when panning methods are used to virtually play them from positions in-between loudspeakers, and paired with visual objects at any position in frontal space. In the past, those studies have often been done with simple laboratory stimuli consisting of signal lights and sounds from single loudspeakers. The SOFE instead allows pairing simulated 3D visual space with 3D audio that includes reverberation and thus depth information.

Of course, many of the ideas offered here can also be useful for laboratories which do not have a true anechoic chamber and so must approximate an open field with echoes reduced by covering the walls with absorptive materials. In that case, experimenters should be cautious about the effects of short-term reverberation that can affect tests of distance perception and interactions between the reflected reflections and those produced by loudspeakers in tests of precedence. Because it is so difficult to predict how such interactions will affect results we would recommend that when possible, the use of anechoic space is preferable.

IV Conclusions

The Simulated Open-Field Environment (SOFE) gives the researcher the opportunity to play sounds over a wide range of levels and frequencies from multiple loudspeakers in an anechoic chamber. Because sounds are played in the free-field, each participant listens with their own ears and individual characteristics of the ears are captured in the sound they hear. This makes an easy and accurate comparison between various listeners with and without hearing devices possible.

Custom calibration software assures each loudspeaker's frequency response is individually equalized and playback software encapsulates the equalization process such that the experimenter plays the sound-as-it-should-be while the system makes sure that it is accurately and identically reproduced from any loudspeaker of the setup.

Room simulation software can create realistic patterns of reflections to one or more sound sources which are played via the loudspeakers inside an anechoic chamber. The natural spatio-temporal pattern of reflections is re-created as each reflection is played from a loudspeaker at the position where it naturally occurs. This assures that binaural hearing processes can use information that is captured in the directional pattern of reflections, e.g. for adaptation to the room response or binaural squelch-type benefits (Djelani and Blauert, 2001).

The sound playback system is complemented by a video reproduction facility which projects images on curtains that cover the loudspeakers. The video system can be used to collect or give feedback from/to the participant or to study auditory-visual interaction.

The SOFE setup is a "Swiss army knife" tool for auditory, spatial hearing and audio-visual research which affords countless possibilities for experiments because of its integration of accurate spatial sound playback with an interactive visual environment. Since the system is driven from multiple computers, it allows for powerful parallel processing of acoustic and visual information while its control from a single Matlab script helps the user to easily interact with it on a high level. Various experiments can thus be created quickly and students can learn to use the setup easily.

Acknowledgments

Abbreviations

EEG	Electro-encephalogram
FIR	Finite-impulse response
HDMI	High Definition Multimedia Interface
HRTF	Head-related transfer function
ILD	Interaural level difference
MLS	Maximum-length sequence
OSC	Open Sound Control
PC	Personal computer
SOFE	Simulated Open Field Environment
SPL	Sound Pressure Level

References

Berkhout AJ, de Vries D, Vogel P. Acoustic control by wave field synthesis. J Acoust Soc Am 1993;93:2764–2778.

Blauert, J. Spatial hearing. MIT Press; Cambridge, USA: 1997.

Bolt RH, Doak PE, Westervelt PJ. Pulse Statistics Analysis of Room Acoustics. J Acoust Soc Am 1950;22:328–340.

Brungart DS, Rabinowitz WM. Auditory localization of nearby sources. Head-related transfer functions. J Acoust Soc Am 1999;106:1465–1479. [PubMed: 10489704]

Deane GB. A three-dimensional analysis of sound propagation in facetted geometries. J Acoust Soc Am 1994;96:2897–2907.

Djelani T, Blauert J. Investigations into the Build-up and Breakdown of the Precedence Effect. Acta Acustica - Acustica 2001;87:253–261.

Gerzon MA. Periphony: With-Height Sound Reproduction. J Audio Eng Soc 1973;21:2–10.

Hafter, E.; Seeber, B. The Simulated Open Field Environment for auditory localization research, Proc. Int. Commission on Acoustics; ICA 2004, 18th Int. Congress on Acoustics; Kyoto, Japan. 4.-9.04.2004; 2004. p. 3751-3754.

ISO, 2003. Acoustics: Determination of sound power levels of noise sources using sound pressure - Precision methods for anechoic and hemi-anechoic rooms (ISO 3745:2003); German version EN ISO 3745:2003, Deutsches Institut fuer Normung and International Standardisation Organization.

Kirszenstein J. An Image Source Computer Model for Room Acoustics Analysis and Electroacoustic Simulation. Applied Acoustics 1984;17:275–290.

Kleiner M, Dalenbäck BI, Stevensson P. Auralization - An Overview. J Audio Eng Soc 1993;41:861–874.

Knowles HS. Artificial acoustical environment control. Acustica 1954;4:80–82.

Krebber, W.; Sottek, R.; Blauert, J.; Brüggen, M.; Dürrer, B. Reproduzierbarkeit von Messungen kopfbezogener Impulsantworten, Fortschritte der Akustik - DAGA '98. DEGA; Oldenburg: 1998. p. 136-137.

Kuttruff, H. Acoustics: An Introduction. Taylor and Francis; London, New York: 2007.

Lee H, Lee BH. An Efficient Algorithm for the Image Model Technique. Applied Acoustics 1988;24:87–115.

Litovsky RY, Colburn HS, Yost WA, Gunzman SJ. The precedence effect. J Acoust Soc Am 1999;106:1633–1654. [PubMed: 10530009]

Moore, BCJ. Cochlear Hearing Loss: Physiological, Psychological and Technical Issues. John Wiley & Sons Ltd; Chichester, UK: 2007.

Pulkki V. Virtual Sound Source Positioning Using Vector Base Amplitude Panning. J Audio Eng Soc 1997;45:456–466.

Rife DD, Vanderkooy J. Transfer-Function Measurement with Maximum-Length-Sequences. J Audio Eng Soc 1989;37:419–444.

Seeber B. A New Method for Localization Studies. Acta Acustica - Acustica 2002;88:446–450.

Seeber B, Baumann U, Fastl H. Localization ability with bimodal hearing aids and bilateral cochlear implants. J Acoust Soc Am 2004;116:1698–1709. [PubMed: 15478437]

Seeber B, Eiler C, Kalluri S, Hafter E, Edwards B. Interaction between stimulus and compression type in precedence situations with hearing aids. J Acoust Soc Am 2008;123:3169.

Seeber B, Fastl H. Localization cues with bilateral cochlear implants. J Acoust Soc Am 2008;123:1030–1042. [PubMed: 18247905]

Seeber, B.; Hafter, E. Perceptual equalization in near-speaker panning. In: Mehra, SR.; Leistner, P., editors. Fortschritte der Akustik - DAGA '07, Dt Ges f Akustik eV. DEGA; Berlin: 2007a. p. 375-376.

Seeber, B.; Hafter, E. Precedence-effect with cochlear implant simulation. In: Kollmeier, B.; Klump, G.; Hohmann, V.; Langemann, U.; Mauermann, M.; Uppenkamp, S.; Verhey, J., editors. Hearing - from Sensory Processing to Perception. Springer; Berlin, Heidelberg: 2007b. p. 475-484.

Seeber, BU.; Fastl, H. In: Brazil, E.; Shinn-Cunningham, B., editors. Subjective Selection of Non-Individual Head-Related Transfer Functions; Proc. 9th Int. Conf. on Aud. Display; Boston, USA: Boston University Publications Prod. Dept; 2003. p. 259-262.

Shelton BR, Searle CL. The influence of vision on the absolute identification of sound-source position. Perception & Psychophysics 1980;28:589–596. [PubMed: 7208275]

Small RH. Closed-Box Loudspeaker Systems Part I: Analysis. J Audio Eng Soc 1972;20:798–808.

Small RH. Closed-Box Loudspeaker Systems Part II: Synthesis. J Audio Eng Soc 1973;21:11–18.

Stecker GC, Hafter ER. Temporal weighting in sound localization. J Acoust Soc Am 2002;112:1046–1057. [PubMed: 12243153]

Vorländer, M. Auralization: Fundamentals of Acoustics, Modelling, Simulation, Algorithms and Acoustic Virtual Reality. Springer; Berlin: 2008.

Wegel RL, Lane CE. The Auditory Masking of One Pure Tone by Another and Its Probable Relation to the Dynamics of the Inner Ear. Physical Review 1924;23:43–56.

Zwicker, E.; Fastl, H. Psychoacoustics, Facts and Models. Springer; Berlin, Heidelberg, New York: 1999.
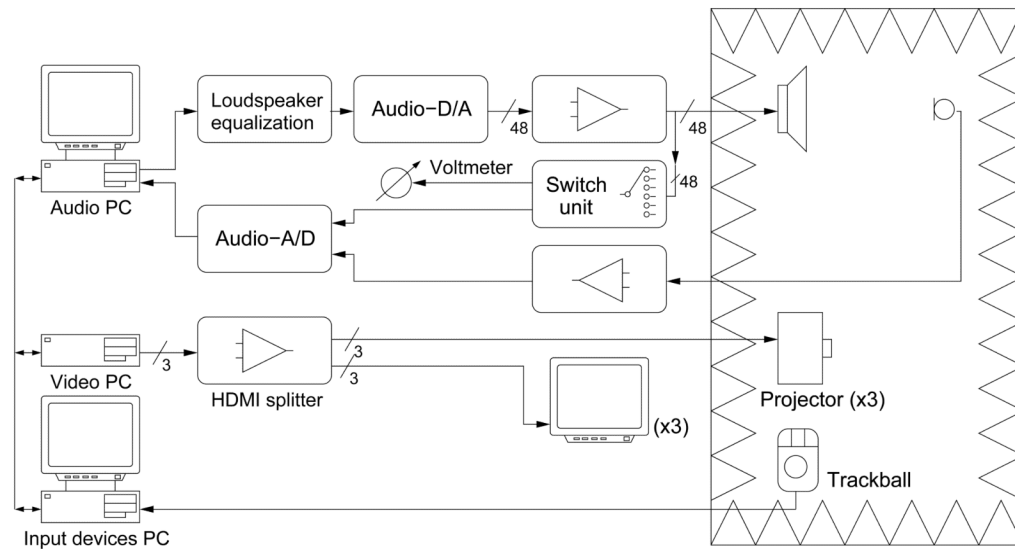
Figure 1.
Overview of the system. The Audio PC handles sound playback and equalization and controls the overall sequence of an experiment. The interactive visual environment is computed on a second computer (Video PC) and displayed via a video projection system on three screens in the anechoic chamber while being monitored in the control-room. A third computer handles input devices like a trackball. Computers communicate via messages on the network. Flexible extensions on additional computers include a motion tracker to record head and body positions and an eye-tracker (not shown).
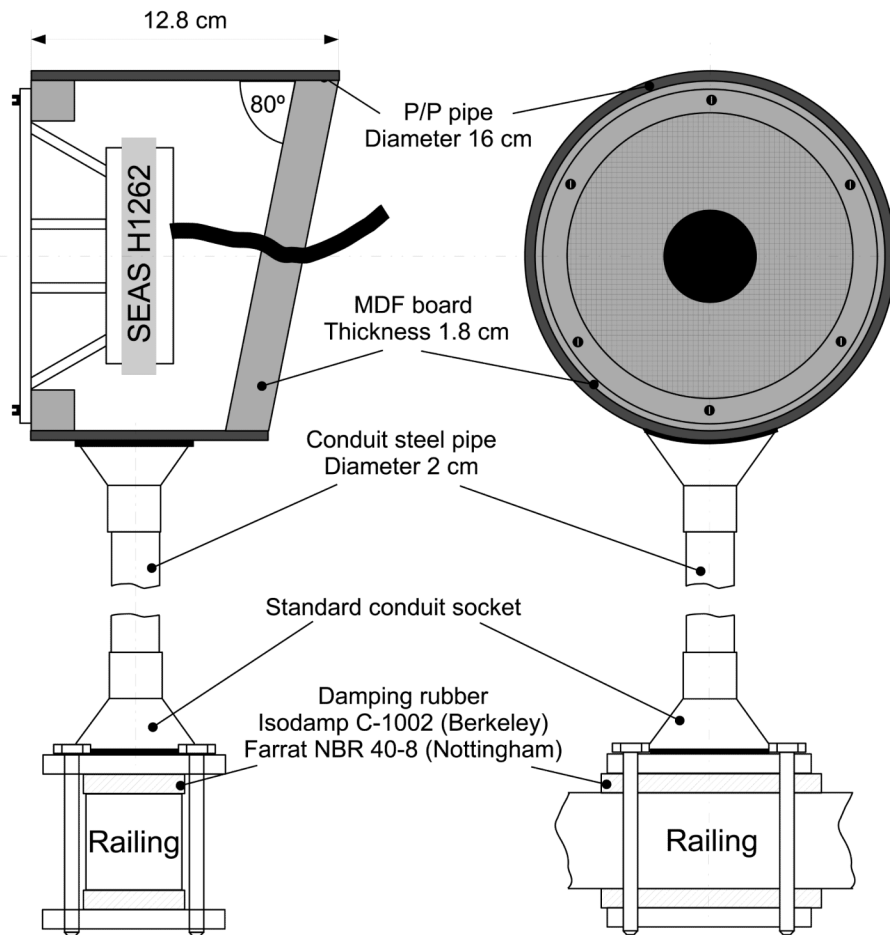
Figure 2.
Loudspeaker system developed for spatial hearing research to minimally affect the sound field in an anechoic chamber while creating high sound pressure level combined with an even frequency response. This is achieved by a cylindrical shape, a slanted back and filling of the housing with damping material. Loudspeakers are mounted on poles and via clamps on a railing that protrudes from the floor of the anechoic chamber. Rubber pads between clamps and the railing minimize pendulum movement and vibration.
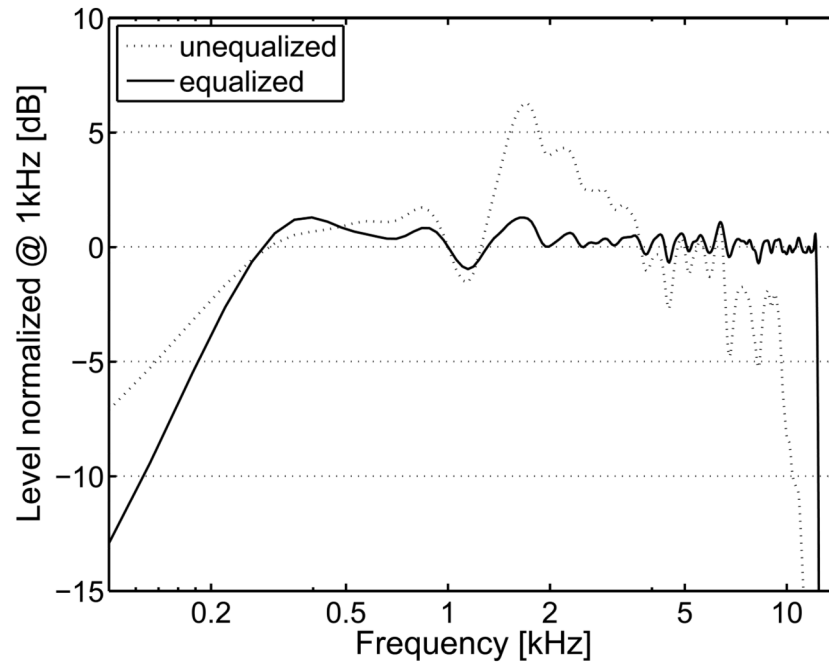
Figure 3.
Frequency response of the custom-made loudspeaker of Figure 2 without (dotted) and with (solid line) digital equalization with a 512-taps FIR-filter. Levels are normalized at 1 kHz to 0 dB.
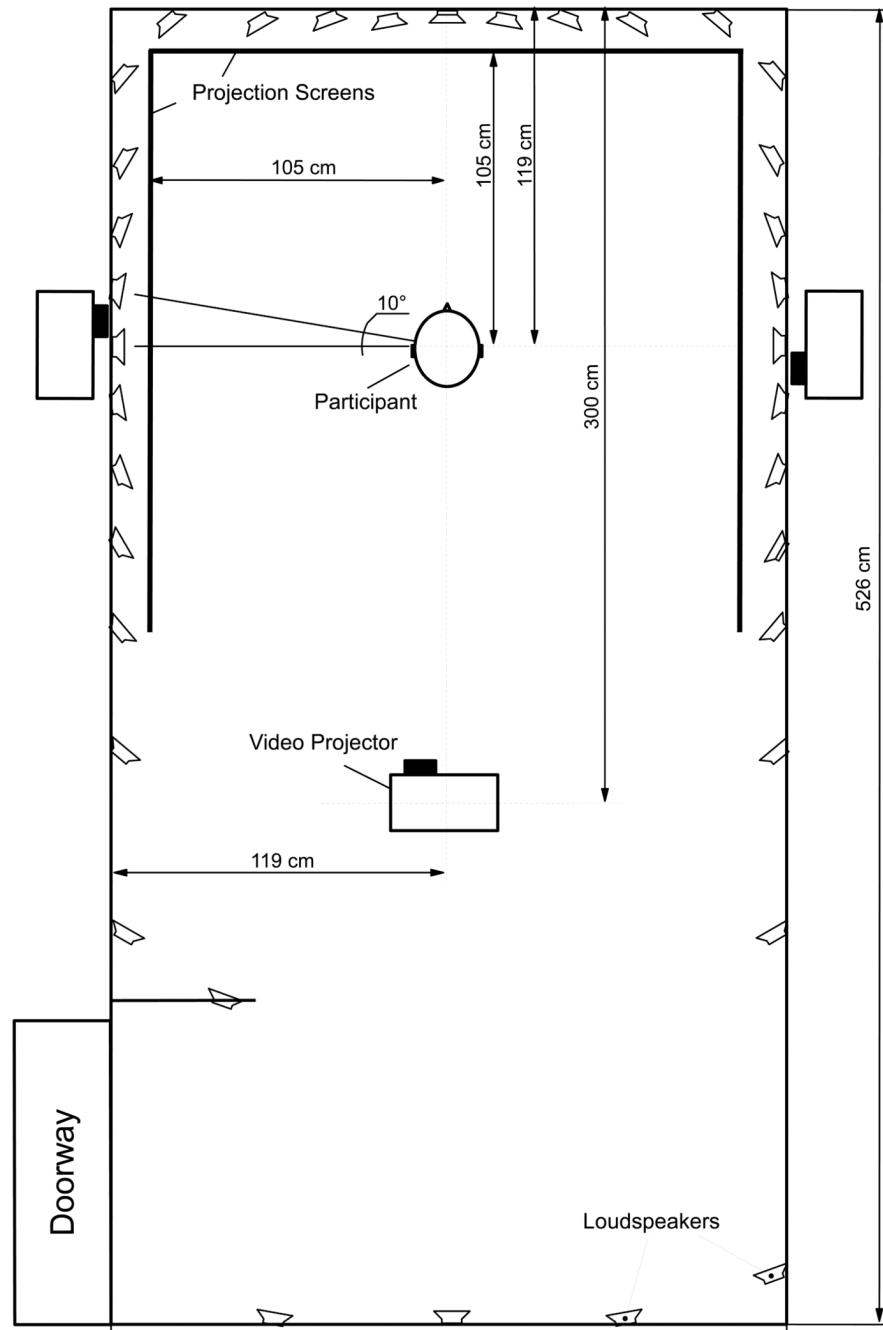
Figure 4.
Sketch of the hardware setup inside the anechoic chamber. Video projectors are mounted at a height of 2.2 m. Loudspeakers placed at elevations other than 0° are not shown.