The Simulated Open Field Environment for Auditory Localization Research

Ervin Hafter and Bernhard U. Seeber

Auditory Perception Lab, Dept. of Psychology, UC Berkeley, Berkeley, CA 94720-1650, USA
Email: hafter@socrates.berkeley.edu

Abstract

Most research on auditory localization has been done in either a non-reverberant environment or using virtual acoustic space presented under headphones. The Simulated Open Field Environment (SOFE) uses computer controlled loudspeakers in an anechoic chamber to create realistic auditory scenes with simulated sources and echoes. Each approach has didsadvantages; for example the anechoic environment represents an unnatural listening condition, while head movements can have an impact on studies with virtual acoustics, and neither offers the subjective experience of immersion in a coherent auditory/visual spatial environment. A major upgrade of the SOFE now features 48 channels of parallel audio in addition to a projected high-resolution visual scene of 80°. Together these provide 360°of acoustic stimuli with a concomitant visual display. Use of such a large number of loudspeakers allows for higher resolution in simulating more complex environments, including multiple sources and reflections as well as auditory motion, and current work is directed toward still greater resolution through amplitude panning between speakers. Distance-dependent damping as well as temporal delay is simulated for all auditory signals. Subjects mark perceived sound azimuths with a visual marker on the screen moved with a trackball. In another task they mark both the direction and distance of a sound by moving virtual visual sources in a 3-dimensional display of visual perspective.

1. Introduction

Originally, our Simulated Open-Field Environment (SOFE) was constructed in an anechoic chamber in order to study the physiology of sound localization through neural recordings in the Inferior Colliculus of cat. Nine loudspeakers spaced at intervals of 20°were placed in front of the anesthetized cat. From these were presented tonal pips of various frequencies, with the responses of individual loudspeakers equalized by potentiometers controlled by a PDP-8L computer. The cat's pinnas were tied in an upright position to control the head-transfer functions (HRTFs) [1].

In its next incarnation the SOFE featured an array of 12 speakers placed in the frontal quadrant of a human listener, and subjects discriminated azimuths with and without simulated echoes. Now, a more elegant computer (PDP-11) presented more complex sounds such as clicks through inverse digital filtering chosen to offset the frequency responses of the individual speakers [2, 3]. As we moved to the PC, the addition of a laser pointer allowed subjects to point at a perceived location and, most recently, that ability allowed for direct measures of binaural precedence, analysis of the effectiveness on localization of each part of an ongoing stimulus, estimation of the relative usefulness of angular, level- and frequency-cues in apparent auditory motion, and tests of the generality of visual bias on auditory localization in an impoverished spatial environment.

The main purpose of current paper is to introduce the once-again reborn SOFE. It allows us to study 360°localization of both laboratory and natural sounds in controllable reverberation, with a new, synchronous wide-angle visual display. This upgrade has required a large reassessment of technology and would not be possible without important improvements in computer speed and software. The intent here is to share some of our experience in building a new SOFE and describe some of the experiments either in planning or underway.

2. Advances of the new SOFE setup

Recent advances in computer technology make it possible to aim for a more accurate spatial reproduction of auditory and visual stimuli in the SOFE. The new hardware allows for 48 parallel channels of audio for speakers surrounding the subject with a maximum of a 7.5°spacing. The improved visual display gives higher resolution and better contrast. The computing requirements for 48 parallel channels of audio synchronized to high resolution video combined with real-time feedback to subjects make the implementation an especially complicated problem which has required a complete rewriting of the software. Because of its ease-of-use for controlling auditory experiments, the programming language Matlab is being used in many laboratories. While we would like to benefit from Matlab in this regard, Matlab does not allow for the necessary parallel processing due to it's sequential structure. In the new implementation, several programs dedicated to specific tasks, e.g. response feedback or visual display, run in parallel on different computers. The different programs are controlled by message exchange from a central experimental rountine residing in Matlab. This new approach keeps the necessary flexibility for experimental control provided by Matlab, but adds true parallel processing as well as the necessary processing power. By relying on message exchange the single programs can be kept simple and modular allowing for an easier and more flexible development.

3. Setup of the SOFE

3.1. Audio and Video-Hardware

A layout of the setup in the anechoic chamber is given in figure 1. 48 loudspeakers (Peerless 981130) are mounted at ear level in an array surrounding the subject seated in the anechoic chamber. Loudspeakers, mounted on individual, vibration damped stalks, can be rotated in order to center their acoustic focus on the head of the subject. Alignment is done by using a laser mounted temporarily on each speaker to illuminate a ping-pong ball suspended in place of a listener. While these speakers are firmly fixed in azimuth, they can, if necessary, be moved to new locations; this can be done, for example, for an experiment where greater angular density is needed in one region. Because the speakers are mounted along the walls of a rectangular anechoic chamber, the distance from speaker to subject is direction dependent. This is of no consequence because the ability to control the level and timing of sounds from each speaker allows for correct simulation of any distance, regardless of the physical position of the loudspeaker. Subjects never see the actual speakers because a flat, visually opaque but acoustically transparent curtain covers the speakers on all four walls.

Given the width of the chamber, seating the subjects 2 m from the front speaker allows for illumination of a visual projection field of $\pm 40°$ (3.4 m wide) horizontally on the front wall. Visual projection is done with a high resolution video projector (1024*768 pixels) with high contrast ratio (>1000:1). Because the wide-angle lenses of the projectors currently available do not allow for an opening $>20°$, the projector has to be mounted behind the subject protruding from the ceiling (c.f. figure 1).

Digital sound signals are delivered from a PC-type computer to 48 channels of external high-quality digital-to-analog converters (2 modules of MOTU Audio 24 I/O, with up to 24 bits resolution, 96 kHz sampling frequency). Analog signals are then amplified by power amplifiers (Crown D-75, 2 channels, each 35 W) before delivered to the speakers.

A custom-made computer-controlled switching unit directs each output of the amplifiers to either a voltmeter, a spectrum analyzer, headphones, or an analog-digital converter. With this the system amplification can be checked automatically through A/D-converters or manually with the voltmeter. Additionally, a measurement mi-

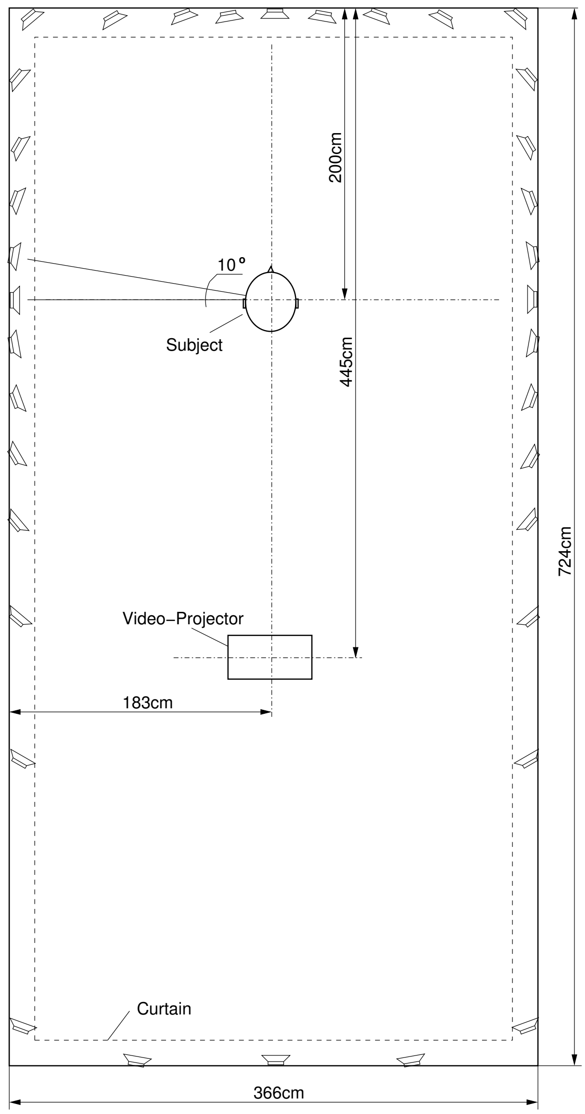

Figure 1: Schematic layout of the setup in the anechoic chamber. This particular configuration utilizes 36 speakers in a 10° spacing relative to the subject.

crophone hanging from the ceiling in the anechoic chamber connects via pre-amplification stages to one of the A/D-converters in the MOTU-system.

The system has been constructed in a way that allows future expansion to more channels. This will allow for inclusion of speakers at different elevations or for increasing the density in the front.

3.2. Measurement and Calibration

Special care is taken to assure the acoustical characteristics of the simulated open field environment. Automated speaker equalization routines have been developed for the Matlab programming environment. For this, a Matlab interface to the 48-channel audio hardware plays maximum-length sequences (MLS) to each speaker while recording the speaker's response synchronously [4]. A recursive algorithm allows conversion of the inverse speaker transfer function into linear-phase FIR filters. Using two uncorrelated MLS-sequences it is possible to determine the temporal delay between speakers at different distances, even if playback and recording are not synchronized. In this way, all of the speakers are equalized at the subject's head position for equal amplitude transfer functions and delays, an important feature for simulation of an open field. Possible changes in the transfer functions of speakers are monitored by daily measurement of them "in place", using a fixed microphone positioned in the center of the anechoic chamber.

3.3. Experimental Control

Experiments are controlled by a central Matlab script. This allows for uncomplicated generation of a variety of auditory stimuli, flexibility in experimental layout and straighforward evaluation of experimental results.

The main computational routine invokes subject data, experimental parameters and calibration routines. It then controls the temporal sequence of events: presentation of auditory stimuli, display of concurrent visual scenes, and collection of subject responses. Because parallel processing can not be done inside of Matlab, most experimental tasks are done in separate programs controlled via TCP/IP message exchange from Matlab. Use of the NTP time protocol permits computers to be synchronized and absolute time stamps to be assigned to experimental events. Parallel processes to the main Matlab routine include the visual display and the generation of visual feedback responses, localization-response methods, and the subject's head position control.

A number of subject-response modes are available for various experimental paradigms. Responses in identification and detection can be collected with a keyboard, a numerical keypad or a custom made response box. In basic localization, subjects use a computer trackball to move a visual marker to the perceived azimuths of auditory signals. For studies of depth perception, the computer projects a 3-dimensional picture of space onto the visual display and the subject can, by moving the trackball in the both the left/right and front/back dimensions indicate perceived distance as well as azimuth. The advantage of a trackball is that by having has no endpoints and being not positioned itself in space it decouples localization from proprioceptive confounds usually present in pointing [5]. A CAD drawing tablet is available for subjects to draw apparent trajectories of sources moving in simulated space.

A head tracker (Polhemus ISOTRACK) continuously monitors the subject's head position throughout the experiment. If the head is out of position, the experiment is suspended while a small display on the screen tells the subject how to place his or her head back into place.

3.4. Visual Display

The presentation of visual scenes by projection opens up a wide range auditory-visual interaction experiments. It can also serve as a feedback device or for the collection of localization responses [5]. The visual display is rendered on a MAC G4 computer. The main visual-display routine is running in MAX/MSP and JITTER, a programming language optimized for the processing of continuous audio and video data. The visual display combines a static, but changeable visual scene, e.g. a room or of loudspeakers in a room, with an positionable object for pointing to localized positions, and with the out-of-position information from the head tracker.

3.5. Simulation of the Open Field

In the Simulated Open Field Environment (SOFE), primary sources and their echos are generated in the open space of the anechoic chamber. This requires a mapping of sources and echos to the discrete positions of speakers which can be done for primary sources and first echos by slight adjustments of the distance or the lateral position of the primary source or the walls involved. In a linear vector-algebra approach, ray vectors relating sources and echoes to the subject are used to calculate the necessary delays and attenuations from relevant speakers for creation of simulated sources and echoes. Another feature of this technique is that different reflective materials can be simulated by modification of the spectral coefficients of the simulated echoic surface. The superposition of temporally and spectrally weighted delta-impulses for each reflection at a single speaker position determines an impulse response for the respective speaker. The computation of the SOFE yields a combination of these individual impulse responses (FIR filters) for all speakers which can be applied to any primary sound source signal.

We plan to examine the feasibility of increasing the spatial speaker resolution through the use of phantom sources and panning techniques. Experiments on the detectability of differences between real sources and phantom sources from the two nearest speakers are underway.

4. Why use the SOFE?

While the SOFE is different from virtual stimulation of space through headphones, both can be useful for studying auditory localization. A major feature of the SOFE is that subjects hear stimuli through precisely their own HRTFs, with no possibility of error due to interpolation [6]. What is more, the head is unencumbered by headphones, presenting a more natural sense of listening in space. For sources from single loudspeakers, the spatial solution remains despite head movements. The large, visual display at a distance across the subject's frontal view also adds the sense of naturalness. Use of 3-D graphics to study auditory distance makes use of the relation between the size of known visual objects and distance.

5. Some experiments currently under development

Three experiments, currently underway or in development, each make special use of the SOFE. In the first, listeners are presented with 1 to N sounds that may be the same or different, all however, within a single one-walled simulated environment. Then a test stimulus is presented for judgment with the wall moved either closer to or farther from the subject. Requiring that they respond "nearer" or "farther" is intended to avoid the common difficulty in interpreting the response "same" with complex stimuli. Of interest is the growth in performance with N and what it tells us about how much the subject learns about a reverberant environment with multiple opportunities.

A second experiment utilizes the ventriloquism aftereffect to study visual bias on sound localization. Of special importance is the relative accuracy of the auditory and visual information when they are combined to make a single percept.

In previous work Seeber and Fastl [7] have shown that high localization accuracy with bilateral cochlear implants (CI) is possible based on the subjects evaluation of interaural level differences (ILD). These patients have no access to the interaural differences of time (ITDs) cues used by normal listeners in complex environments. The plan is to use the SOFE to ascertain how well CI-listeners can do in reverberant enviroments. .

6. References

[1] A. L. Leiman and E. R. Hafter, "Responses of inferior colliculus neurons to free field auditory stimuli," *Exp. Neurol.*, no. 35, pp. 431–449, 1972.

[2] E. R. Hafter and F. Briolle, "Localization of complex sounds in the presence of echoes," in *Proc. of the 13th Int. Congress on Acoustics*, 1989, pp. 495–498.

[3] E. R. Hafter, K. Saberi, E. R. Jensen, and F. Briolle, *Localization in an Echoic Environment*. Pergamon, Oxford, 1992, pp. 555–561.

[4] D. Rife and J. Vanderkooy, "Transfer-function measurement with maximum-length-sequences," *J. Audio Eng. Soc.*, vol. 37, no. 6, pp. 419–444, 1989.

[5] B. Seeber, "A new method for localization studies," *Acta Acustica – Acustica*, vol. 88, no. 3, pp. 446–450, 2002.

[6] B. Seeber and H. Fastl, "Subjective selection of non-individual head-related transfer functions," in *Proc. 9th Int. Conf. on Aud. Display*, E. Brazil and B. Shinn-Cunningham, Eds. Boston, USA: Boston University Publications Prod. Dept., 2003, pp. 259–262.

[7] B. Seeber, H. Fastl, and U. Baumann, "Mechanismen der Lokalisation mit bilateralem Cochlea Implantat (Mechanisms of localization with bilateral cochlear implant)," in *Fortschritte der Akustik – DAGA 2003*. Oldenburg: DEGA, 2003, accepted.

This work is supported by NIH Grant NIDCD 00087.