

See discussions, stats, and author profiles for this publication at: <https://www.researchgate.net/publication/261871651>

Design and verification of HeadZap, A semiautomated HRIR measurement system

Conference Paper · May 2006

CITATIONS

16

READS

282

6 authors, including:



Durand R Begault

NASA

97 PUBLICATIONS 2,612 CITATIONS

[SEE PROFILE](#)



Martine Godfroy-Cooper

San Jose State University/ U.S. Army DEVCOM AvMC TDD

49 PUBLICATIONS 222 CITATIONS

[SEE PROFILE](#)



Joel D Miller

NASA

26 PUBLICATIONS 291 CITATIONS

[SEE PROFILE](#)



Agnieszka Roginska

New York University

82 PUBLICATIONS 410 CITATIONS

[SEE PROFILE](#)

Some of the authors of this publication are also working on these related projects:



Spatial Auditory Cueing for helicopters in Degraded Visual Environment [View project](#)



HoloDeck [View project](#)



Audio Engineering Society

Convention Paper 6655

Presented at the 120th Convention
2006 May 20–23 Paris, France

This convention paper has been reproduced from the author's advance manuscript, without editing, corrections, or consideration by the Review Board. The AES takes no responsibility for the contents. Additional papers may be obtained by sending request and remittance to Audio Engineering Society, 60 East 42nd Street, New York, New York 10165-2520, USA; also see www.aes.org. All rights reserved. Reproduction of this paper, or any portion thereof, is not permitted without direct permission from the Journal of the Audio Engineering Society.

Design and verification of HeadZap, a semi-automated HRIR measurement system

Durand R. Begault¹, Martine Godfroy², Joel D. Miller³, Agnieszka Roginska⁴,
Mark R. Anderson⁵ and Elizabeth M. Wenzel¹

¹ Human Systems Integration Division, NASA Ames Research Center, Moffett Field, CA 94035 USA
Durand.R.Begault@nasa.gov

² San José State University Foundation, NASA Ames Research Center, Moffett Field, CA 94035 USA
mgodfroy@mail.arc.nasa.gov

³ VR Sonic Incorporated, San Francisco, CA USA
jdmiller@vrsonic.com

⁴ AuSIM Incorporated, Palo Alto, CA 94306 USA
aroginska@ausim3D.com

⁵ QSS Group, NASA Ames Research Center, Moffett Field, CA 94035 USA
manderson@mail.arc.nasa.gov

ABSTRACT

This paper describes the design, development and acoustic verification of HeadZap, a semi-automated system for measuring head-related impulse responses (HRIRs) designed by AuSIM incorporated and modified by the NASA Ames Research Center Spatial Auditory Display Laboratory. HeadZap utilizes an array of twelve loudspeakers in order to measure 432 HRIRs at 10° intervals in both azimuth and elevation in a non-anechoic environment. Applications to real-time rendering using SLAB software and an audio-visual localization experiment are discussed.

1. INTRODUCTION

1.1. Overview

This paper describes the HeadZap system for measurement of Head-Related Impulse Responses (HRIRs). The time-domain HRIRs and their frequency-domain equivalent HRTFs (Head-Related Transfer Functions) are used to capture spatial hearing cues for a specific measured person. The cues can be recreated via virtual acoustic rendering and/or for psychoacoustic research and post-measurement analysis.

This paper first describes the features of the HeadZap system. Second, the system is evaluated –“verified”– in terms of its accuracy and repeatability of the acoustical measurements. Third, system verification from the standpoint of human HRTF measurement repeatability is shown. Finally, a psychoacoustic verification is presented that examines the relationship between real and virtual source location, including application within an ongoing study by author Martine Godfroy. We conclude that the system measurements are repeatable and verifiable within reasonable limits.

1.2. Motivation for development of the system

The Spatial Auditory Display Laboratory at NASA Ames Research Center is currently a part of the Advanced Controls and Displays group within the Human Systems Integration Division. Our mission is to develop improved auditory interfaces for aeronautic and space applications, including virtual environment simulation for training, communication displays, caution and warning systems, orientation-situational awareness signaling, and advanced concepts.

Since 1987, we have had considerable experience with HRTF measurement systems. Previous systems used in collaborative work with the University of Wisconsin were very time-consuming and required an anechoic chamber [1]. Later, an innovative device called SNAPSHOT was developed and implemented in our laboratory, principally through the work of Jonathan Abel (then at Crystal River Engineering). This system used a single loudspeaker and was capable of measuring a rough grid of 30° azimuth intervals at six elevations in a reflective environment. The measurement technique used microphones positioned in the blocked-meatus with Golay codes as the probe stimulus. In contrast to

previous HRTF measurement systems, the listener moved on a rotating stool relative to a fixed loudspeaker position in order to measure different azimuth positions; the loudspeaker was then repositioned for different elevation measurements. The system was limited to a single diffuse-field equalization based on a specific set of headphones.

In 1998, we began work with William Chapin of AuSIM Incorporated to make major improvements to the SNAPSHOT system, while keeping its best features. The system specifications for the improved system can be summarized as follows:

- Increase the density of the measurement grid from 72 to 432 measurements on a sphere about the listener.
- Faster means of subject positioning while maintaining reliability and repeatability of measurement.
- Improve fidelity in the time domain and optimize microphone signal-to-noise ratio (SNR).
- Improve method for equalization of headphone response; allow for different headphone equalizations; implement equalization to the free-field.
- Develop tools to easily adapt measured HRIRs to the PC-based rendering system used in our laboratory (the SLAB system – see [2-4]).

Each of these desired specifications were addressed in the overall system design for both software and hardware. We begin with an overview of the signal measurement path, followed by a description of the loudspeaker setup and “closed-loop” subject orientation system used to increase density of measurements and reduce measurement time. Next, we briefly review the hardware used to upgrade the fidelity of the system. Finally, the software tools used for signal equalization and for adapting the measurements to the real-time rendering system are described.

1.3. Overview of the signal measurement path

The flow of the signal measurement path of HeadZap has features in common with many HRIR measurement systems (Figure 1). The first step involves measurement of raw HRIRs using miniature microphones at the

blocked-meatus position of both ears of the listener. The second step involves applying a time window to the measured response, so that reflections from nearby surfaces and the propagation time of the signal are eliminated. The third step involves equalizing the signals such that the specific measurement loudspeaker and microphone transfer functions are compensated. Fourth, the headphones to be used for playback are measured *in situ* on the listener or on a mannequin head, so that compensatory equalization can be applied to the ensemble of measurements (referred to as an ‘HRTF map’). Finally, the HRTF map is processed for use in the spatial rendering system, including minimum phase processing, interaural time delay calculation, and other factors, as detailed below in sections 1.7-1.10.

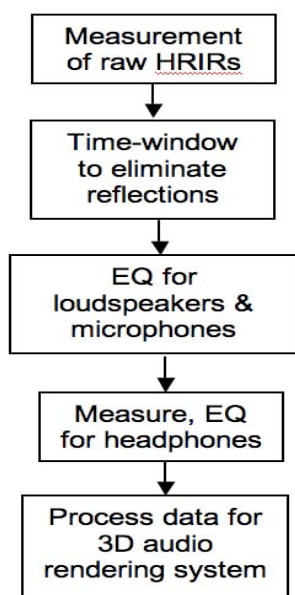


Figure 1. Overview of HeadZap algorithm.

1.4. Loudspeaker system design

In order to increase the density of measurements while simultaneously decreasing the measurement time, multiple loudspeakers are mounted asymmetrically about the listener’s measurement position. As with the previous SNAPSHOT system, the listener moves with respect to the fixed loudspeaker positions on a rotating stool. For each azimuth orientation of the listener, HRIR measurements at twelve different elevations can then be taken in rapid succession before the next orientation of the listener. The process is repeated until a total of 432 measurements have been made.

Twelve loudspeakers (AuSIM “AuProbe” AuPBE101, 63.5 mm diameter driver) were installed in the measurement area, a double-walled soundproof booth (3.0 x 3.0 x 2.4 m dimension, galvanized perforated steel inner surface over absorptive material; Industrial Acoustics Company). Refer to Figure 2 for a schematic of the arrangement and Figure 3 for a photo of the system). Loudspeakers are located at a distance of 0.9m from the center of the speaker sphere. The 0° elevation speaker is located at a height of 1.235m above the floor. The directivity of the loudspeakers is indicated in Figure 4; the frequency response of each loudspeaker was individually compensated (see section 1.8).

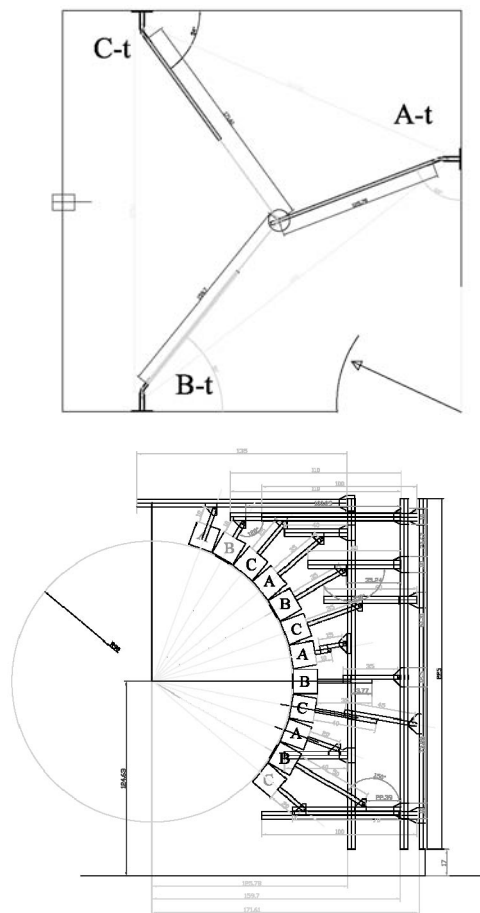


Figure 2. Loudspeaker arrangement. Top: plan view; Bottom: section view with poles A-B-C overlapped.



Figure 3. Fisheye lens view of loudspeakers within booth. Transmitter and microphone inputs are mounted on a wooden pole descending from the ceiling at the upper right of the head; aiming devices (section 1.5) are distributed at eye height.

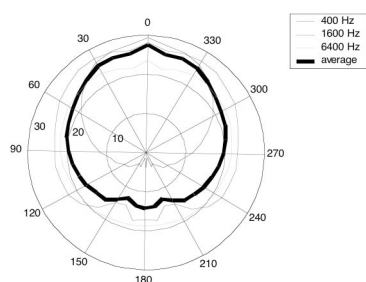


Figure 4. Directivity of loudspeakers used.

Speakers were installed at elevations ranging from $+70^\circ$ to -40° , with 10° intervals (lower part of Figure 2). Each speaker was mounted on an aluminum arm, which was attached to an aluminum post. A total of four speakers were positioned on each post, with a 30° interval between the speakers. The three posts were mounted onto the walls of the measurement room. The configuration resulted in no single loudspeaker directly

facing another one, thereby avoiding a reflective ‘cage’ effect. The coordinate system used here is a single-pole system where the nominal position directly in front of the listener is defined as 0° azimuth, 0° elevation. Relative azimuths to the listener’s right are positive, azimuths to the listener’s left are negative, and directly behind is defined as 180° azimuth. Elevations above 0° are defined as positive 0° to 90° and below 0° are defined as 0° to -90° .

1.5. “Aimer”: interactive subject positioning system

The requirement for reducing measurement time has both obvious and non-obvious benefits. Our previous system SNAPSHOT involved noting the output of a head tracker on a screen by the experimenter; he/she would verbally call out to the subject to make positional adjustments, and then have the subject ‘freeze’ to make the measurement ‘snapshot.’ The current HeadZap system avoids this problem through the use of an interactive, ‘closed-loop’ feedback system using signal lamp feedback.

The HeadZap system incorporates a six-degree-of-freedom head tracking system that is attached to the head. An electromagnetic tracker (Polhemus FastTrak) is installed in the measurement booth and used during HRTF measurements (ref. Figure 3 and Figure 5). This 6 degree-of-freedom tracker has a latency of 4ms and an update rate of 120 Hz when one sensor is connected. For multiple sensors, the update rate is shared between sensors. The XYZ static resolution is 0.03 inches RMS, and the resolution for orientation is 0.15°. The maximum range between the transmitter and receiver is 1.52 m. The head tracker is connected to the host computer via an RS-232 interface. The transmitter is attached to a wooden beam, which is connected to the ceiling of the isolation booth. The sensor is placed on top of the subject's head during the measurement.



Figure 5. Microphones are placed near the entrance of the ear canal in a blocked-meatus position. The tracker sensor is held to the head via a Velcro strap.

The head tracking system allows the subject's positional data to be recorded during the measurement process. The tracker data are also passed in real time to a system of visual targeting devices, which allows the subject being measured to easily orient his/her head during the process.

The visual targeting system consists of an array of panels distributed around the subject. Each panel consists of thirty red light emitting diodes (LEDs) and one green LED. The arrangement of the LEDs is shown in Figure 6. The system works by providing intuitive feedback to the subject about his/her head position. Taking azimuth as an example, if the subject is five or

more degrees left of the target, all five azimuths left LEDs will be illuminated. As the subject closes to within one degree (plus or minus) of the desired azimuth, the LEDs will extinguish. Three red LEDs would indicate to the subject that he/she is within three degrees of the desired position and so on. Pitch and roll work in a similar fashion. Once the subject has centered him/herself to within one degree in all three axes, the green LED illuminates and the measurement is taken.

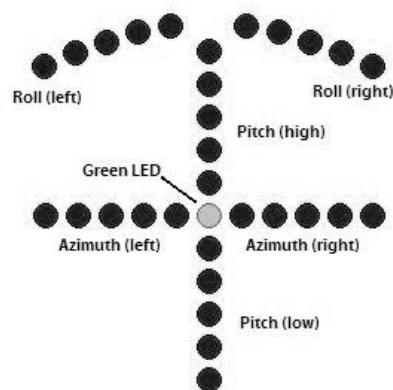


Figure 6. Lamp indicators for subject feedback.

The visual targeting system only allows the subject to maintain his/her head orientation (pitch, roll, yaw) relative to the speaker array; however, it cannot control the subject's physical position (x, y, z) within the center of the sphere created by the speaker array. In order to ensure that the subject's head stays centered within the sphere, a special stool was constructed that incorporates a chin rest. The stool rotates along the vertical axis of the sphere, and the chin rest ensures that the ears are at the proper height and centered along the vertical axis. The chin rest keeps the subject from leaning in any direction, thus controlling position in the horizontal plane. If the subject becomes fatigued and attempts to slouch or straighten his/her back, the head will pitch accordingly and the error is shown on the visual targeting system, thus controlling the vertical position. (We found no significant acoustical influence of the chin rest on the actual HRIR measurements).

1.6. Hardware improvements for upgrading fidelity

Below we briefly review some hardware improvements that enabled better time resolution and SNR. In some cases this involved taking advantage of optimal off-the-

shelf hardware. Figure 7 indicates the signal flow through the hardware system.

One significant hardware improvement was the decision to implement the measurement system with a 96 kHz sampling rate. While such a high sampling rate may provide improvement in spectral frequency resolution, it is probably not necessary for synthesis in the audible frequency range. However, a case can be made for a high sampling rate when considering human perceptual limits for the Interaural Time Difference (ITD). For example, numerous studies of lateralization have shown discrimination thresholds for the ITD on the order of 6 to 10 microseconds, depending upon the spectral content and duration of the stimulus [5-8]. A 96 kHz sampling rate provides a temporal sampling resolution of ~10 microseconds, thus approximately matching the known limits of the human auditory system.

A Multichannel Audio Digital Interface (MADI; RME ADI-648) supports 24-bit analog-to-digital and digital-to-analog conversion with up to 64 channels at 48 kHz and 32 channels at 96 kHz. The MADI is equipped with technologies that allow for synchronization between all channels and offer notification of error states (SyncAlign®, SyncCheck®). The analog-digital and digital-analog converters are connected to the computer via fiber optic cables.

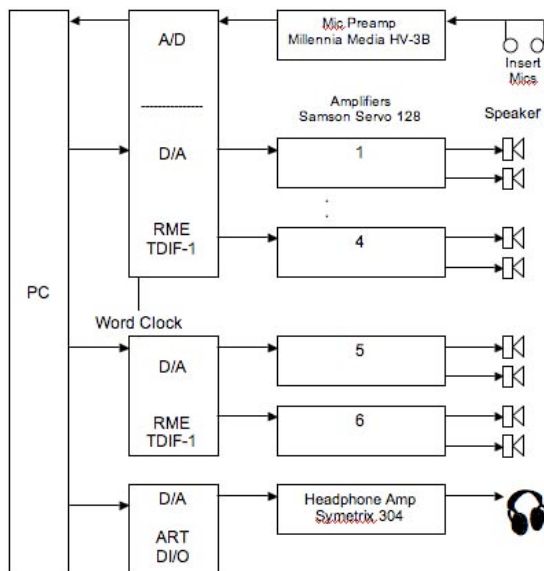


Figure 7. Diagram of the input / output signal paths.

A low-noise microphone preamplifier (Millenia HV-3B) is used for acoustic measurements. The HV-3B is a two-

channel, stereo-matched microphone preamplifier. It combines a matched octet of discrete bipolar transistors with a laser-trimmed DC coupled FET-based output driver stage. The HV-3B is an entirely "double-balanced," wide dynamic-range stereo microphone preamplifier.

A total of six stereo power amplifiers (Samson Servo 120) were installed to power the twelve loudspeakers used for acoustic measurements. The Servo 120 has 60 watts of power per channel, with independent controls for each channel with 41-step increments. Its specifications indicate a flat frequency response from 10 Hz to 100 kHz.

1.7. Impulse response measurement and signal processing for time delay estimation and minimum phase

To measure the directionally dependent impulse responses, a test signal representative of the ideal impulse is used. The measurement system uses Golay codes as the test signal to measure HRTFs. Golay codes [9] are a series of two sequences $a(t)$ and $b(t)$ having complementary autocorrelation side lobes so that their autocorrelations sum to an impulse,

$$a(t) \bullet a(t) + b(t) \bullet b(t) = \delta^2(t) \quad (1.1)$$

where δ^2 represents the combined energy of the $a(t)$ and $b(t)$ sequences.

A system characterized by the impulse response $h(t)$ will return the convolution of the system response with the probe sequence: $a(t) \ast h(t)$. Similarly, the response of the system to the sequence $b(t)$ is $b(t) \ast h(t)$. Correlating the $a(t)$ sequence response with $a(t)$ gives

$$a(t) \bullet [a(t) \ast h(t)] = [a(t) \bullet a(t)] \ast h(t) \quad (1.2)$$

Similarly, correlating the $b(t)$ sequence response with $b(t)$ yields

$$b(t) \bullet [b(t) \ast h(t)] = [b(t) \bullet b(t)] \ast h(t) \quad (1.3)$$

Adding together the correlated sequence responses, and normalizing by the Golay code sequence energy δ^2 , leaves the system impulse response $h(t)$,

$$\begin{aligned} & 1/\delta^2 [a(t) \bullet a(t) + b(t) \bullet b(t)] \ast h(t) \\ & = \delta(t) \ast h(t) = h(t) \end{aligned} \quad (1.4)$$

A complementary Golay sequence can be constructed by a “negate and concatenate” algorithm. If a and b are the L -digit binary representation of a complementary pair, then a new pair of complementary codes a_l and b_l , of length $2L$, can be created by concatenating b to a , and concatenating $-b$ to a .

In additive noise, if $n_a(t)$ were present when measuring the system response to the $a(t)$ sequence, and if additive noise $n_b(t)$ were present when measuring the response to the $b(t)$ sequence, then the system impulse response estimate above becomes

$$h(t) + [a(t) \cdot n_a(t) + b(t) \cdot n_b(t)] / \square^2. \quad (1.5)$$

Assuming that $a(t)$ and $b(t)$ are uncorrelated with their respective noise, the SNR gain is seen to be large if $a(t)$ and $b(t)$ are spread out over time. In this way, the denominator term $\square^2$ representing the total sequence energy is made large, and the numerator term, roughly the product of the total noise energy and the mean absolute $a(t)$ and $b(t)$ sample values, is held in check.

When $a(t)$ and $b(t)$ are bi-level sequences taking on values ± 1 , the SNR gain becomes maximized. For such length L sequences, the total energy is $2L$ while the mean absolute value is only one. In the presence of noise sequences comprised of uncorrelated samples, it follows from (1.5) that the noise appearing in the measured impulse response has on average $1/(2L)$ of the power of the noise sequences themselves. The SNR gain of a Golay code impulse response measurement system using L -sample long bi-level sequences is expressed in units of dB, $10\log_{10}(2L)$.

When using a series of two sequences to measure the impulse response, it is important to consider the effect of the potential response remainder from the first sequence in the second. Thus, before probing the system with the second sequence, we must wait until the signal has sufficiently decayed. HeadZap uses T_{60} estimation at the beginning of the measurement process to determine the wait time between the $a(t)$ and $b(t)$ probe sequences. Our implementation uses 2048 samples at 96 KHz and a 0.2 sec sequence interval.

The measured impulse responses are processed and stored as minimum phase filters with the interaural time differences stored separately. Before converting the filters to minimum phase, the interaural time differences are extracted. To obtain the time delay, the onset of the left and right impulse responses is detected by using an

energy-based onset detection algorithm. The left and right channels are scanned independently to determine the energy at each point. The onset of the impulse response is computed by calculating the point at which the energy of the signal has reached at least 33% of the signal’s maximum energy value. This time is used at the onset of the signal. Using the onset time of the left and right responses, the difference between the two is computed to obtain the resulting interaural time difference between the two ears.

Following the ITD extraction process, the filters are converted to minimum phase. The minimum phase of the filter is obtained by taking the Hilbert transform of the logarithm of the spectral density of the impulse response.

1.8. Signal processing for equalization 1: loudspeaker – microphone compensation

The raw measured impulse response contains within the measured signal the response of the equipment used for the measurement. The equipment characteristics include the speaker and microphone frequency response (see Figures 8-9), the analog-to-digital and digital-to-analog converters, sound card, speaker amplifier and microphone pre-amplifier responses. To fully extract the binaural filters, the measured responses must be compensated for the measurement equipment.

The measured transfer function can be defined as:

$$Y_s(\square) = X_s(\square) S(\square) M(\square) H(\square) \quad (1.6)$$

where $X_s(\square)$ is the test signal, $S(\square)$ is the transfer function of the loudspeaker and amplifier, $M(\square)$ is the transfer function of the microphone and pre-amplifier, and $H(\square)$ is the head-related transfer function.

The free-field transfer function is used as the compensation transfer function to the system, defined as:

$$Y_{\text{comp}}(\square) = X_s(\square) S(\square) M(\square) \quad (1.7)$$

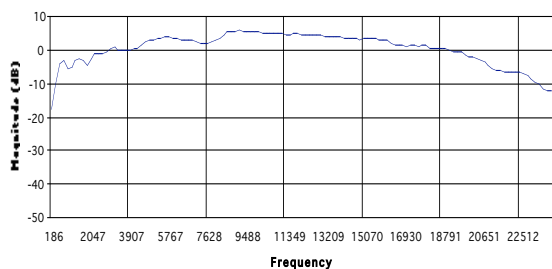


Figure 8. Typical loudspeaker response.

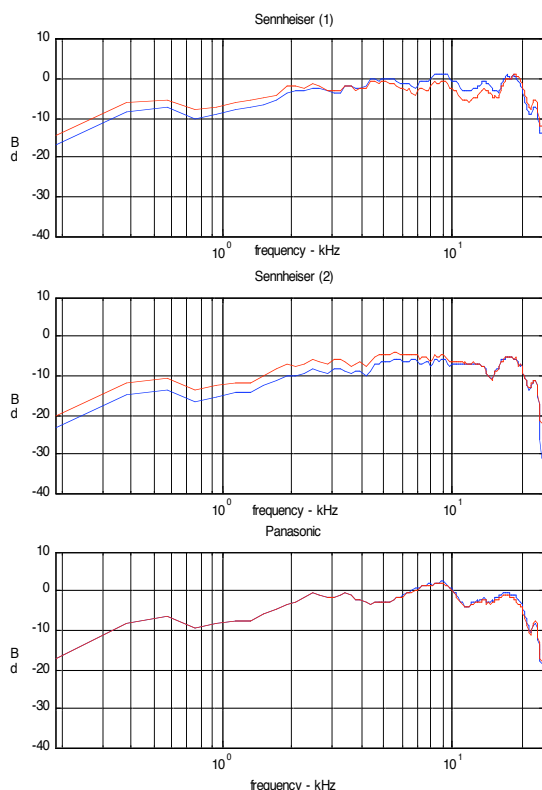


Figure 9. Sennheiser KE4-211-2 (top, middle) and Panasonic microphone response (bottom).

The transfer function for the measurement apparatus is obtained by measuring the free-field response of the closed-loop system. In other words, the free-field response is measured by placing the measurement microphones at the center location (where the center of the head would be during HRTF measurements), playing the test signal through the loudspeaker and measuring the resulting impulse response for the left and right channel microphones. The inverse transfer function is used to equalize the measurements. The

HRTF ($H(f)$) is obtained by removing the free-field transfer function from the measured response:

$$H(f) = Y_s(f) / Y_{comp}(f) \quad (1.8)$$

The setup described in this paper contains a total of twelve loudspeakers that are used for measuring impulse responses. Although the speakers were designed to contain equivalent characteristics, small differences between the responses of the speakers exist. Thus, a free-field equalization filter is measured for each loudspeaker.

1.9. Signal processing for equalization 2: headphone compensation

When sounds are reproduced over headphones, an added coloration is introduced: the spectral coloration of the headphones. To eliminate this effect, we have devised a system that measures the headphone characteristic and removes this from the measured HRTFs.

The headphone equalization measurement process is similar to the HRTF measurement process. Headphone compensation may be measured on a human subject, or a mannequin. The subject is outfitted with a pair of microphones placed at the entrance to the ear canal. Headphones are positioned by the subject on top of his/her ears. A measurement follows. Due to the fact that the subject places a set of headphones differently each time, three measurements are taken. The subject is asked to remove and replace the headphones each time. Thus, the measurements provide a more accurate estimate of the headphones compensation.

The compensation filter is obtained by the following process. First, an average characteristic of the headphones is computed by taking the mean of the three measured responses. Second, the response is smoothed by using a 1/3 octave band smoothing. It has been observed that different placements of the headphones over the ears often result in a movement of the sharp spectral characteristics (notches and peaks) on the frequency axis. The smoothing is performed in order to remove these sharp notches. Next, the signal is band-passed between 400 Hz and 17 kHz and the inverse filter is calculated for the band-passed signal. Last, the filters are converted to minimum phase.

1.10. Processing of measurements for real-time rendering

In practice, the measured HeadZap HRTF database is rendered in real-time using the virtual acoustic environment rendering system SLAB [4]. SLAB was developed by NASA for conducting psychoacoustic experiments and prototyping spatial auditory displays and communication systems. A few post-processing steps are required in order to use a HeadZap HRTF database with SLAB. All post-processing is performed in double-precision floating-point using a specialized MATLAB application (AuSIM AHMTools). The SLAB databases are saved and processed in real-time using single-precision floating-point.

Because SLAB is a real-time system, it is important to keep runtime computational requirements as low as possible. Thus, the HeadZap HRTF databases are down-sampled from 96 kHz to 44.1 kHz. At this sample rate, an HRIR length of 128 taps suffices to capture the response. Another performance optimization results in a requirement that the SLAB database grid be uniformly sampled about the measurement sphere. The density of the grid is variable, but typically azimuth and elevation increments of 10° are used. Where needed, grid resampling is performed using biharmonic spline interpolation [10], an interpolation technique well suited to spherically distributed data. The resultant database can then be used with SLAB to render a virtual sound source anywhere in auditory space.

Once listener-relative azimuth and elevation values have been calculated for an arbitrary sound source and listener location, the relevant HRIR and ITD are formed by interpolating up to four HRTF database locations. Since the HRIRs are minimum phase, this results in a simple weighted linear interpolation of HRIR filter taps. To simulate sound path arrival time behavior, a sound source is attached to a fractionally indexed delay line. The ITD value is used as a left-ear/right-ear offset into the delay line to model a frequency-independent ITD. Once left-ear and right-ear sound samples are extracted from the delay line they are convolved with the left-ear and right-ear HRIR filters, respectively. The result of this operation is a virtual sound source tailored for headphone display. The SLAB spatial rendering algorithm is discussed in much greater detail elsewhere [2,3].

2. SYSTEM ACOUSTIC VERIFICATION

2.1. Measurement of free-field transfer function

In order to verify that the system equalization and time windowing described in section 1 was performing within expected tolerances, a set of measurements were made without the presence of a head so as to initially remove variables such as movement, variation in microphone placement, and inter-subject variability. The acoustic response of the system was measured at the central position (i.e. the center of where the subject's head would be). Measurements focused on room reflections, system interactivity, and head tracking noise interference.

Measurements of the room reflections are depicted in Figure 10. These measurements were sampled at 96 kHz. From this Figure, it can be seen that the first of the wall reflections occurs at sample 214 (speaker elevation $+70^\circ$). These measurements were taken at the central position.

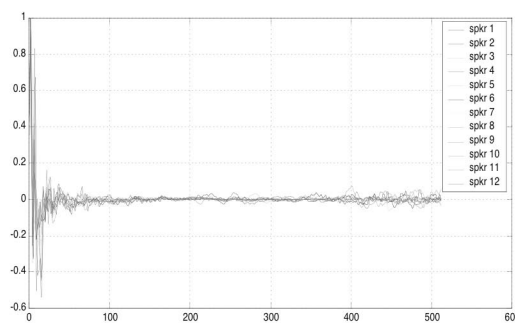


Figure 10. Free-field acoustic response of all speakers at the central position (96kHz).

To calculate the maximum filter length possible, we must take into account the distance offset from the central position to the location of the ears. Assuming a maximum head size of .85msec (82 samples at 96kHz), the maximum filter length should not exceed 173 samples. This indicates that the filter should not exceed 1.8msec in length. At 44.1kHz sampling rate, the filter length should not exceed 79 samples.

2.2. System interdependency

A system containing multiple speakers organized in a spherical setting may result in an interaction between the speakers. To test whether a speaker interaction was occurring in the current setting, measurements were performed with and without all speakers present. Microphones were placed in the free-field at the center of the speaker sphere. First, a measurement was taken using a speaker with all other speakers present in the room. Next, all speakers, except for the measuring speaker, were removed.

The top plot of Figure 11 shows the frequency response of the measurement of the 0° elevation speaker with and without all other speakers present. The bottom portion of Figure 11 displays the difference in the responses between the two measurements. The effect of other loudspeakers can be seen to be negligible.

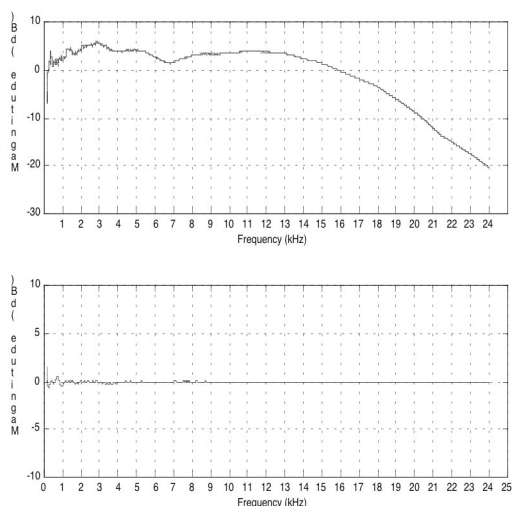


Figure 11. Response for the 0° elevation speaker measured in the free-field (see text).

2.3. Head tracker interference noise measurement

Due to possible interference caused by the magnetic tracking device, measurements were made to test the amount of interference in the signal when a head-tracking device is used during measurements. Two sets of measurements were taken using a dummy head with blocked-meatus microphones. The first set of measurements was recorded with the subject not

wearing the head-tracking device and the tracker turned off. For the second measurement, the tracker sensor was carefully placed on the top of the subject's head. Only measurements of the frontal position were taken. Results indicate no significant interference in the measurements taken with the sensor affixed on the top of the subject's head and the tracking device turned on (see Figure 12).

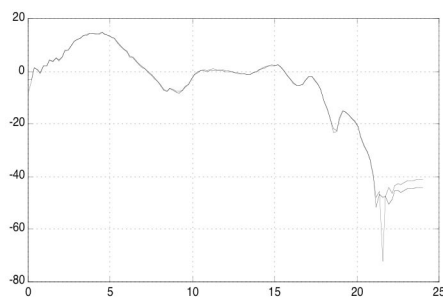


Figure 12. Head-tracking interference measurement. Left-ear frontal position measurement on B&K HATS with and without tracker active. The x-axis indicates frequency in kHz; the y-axis indicates relative amplitude in decibels.

2.4. Centering

The center of the speaker sphere was measured acoustically based on the signal's time of arrival. A single Panasonic WM-61 microphone was placed at the center of the sphere. Golay codes were used as the test signal. Measurements were taken at 96kHz sampling rate. Results are shown in Figures 13-14. Acoustic measurement results show a centering placement to within 1 sample at 96kHz corresponding to 3.6mm accuracy.

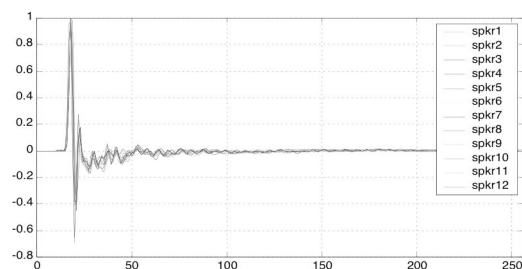


Figure 13. Acoustic centering measurement. The x-axis indicates time in samples, the y-axis indicates relative amplitude.

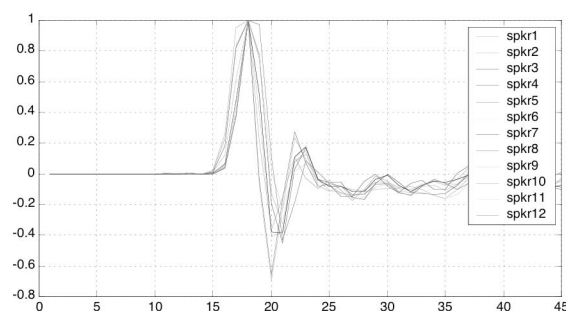


Figure 14. Acoustic centering measurement (zoom-in on x-axis of Figure 13, first 45 samples).

3. HEAD RELATED TRANSFER FUNCTION (HRTF) REPEATABILITY

In this paper, the term HRTF *variability* relates to inter-individual differences, i.e., to the highly individualized “signatures” observed between different listeners’ HRTFs. Intra-individual variability has previously been described as measurement error (e.g., [11]). Here, we refer to this parameter as HRTF *repeatability* or the deviation in measurements taken from a single subject, and produce a quantitative estimate of the reliability of the system and procedure. Indeed, HRTF measurements are susceptible to various biases, such as inaccuracy in positioning the test subject, inaccuracy in positioning the microphones within the ear canals or even small variations in the SNR of the probe signals. We analyzed the variability between three non-consecutive (> one hour interval) measurements that required repositioning of the subject and the microphones for two of the three subjects measured. The specific contribution of the headphone-to-ear measurements is assessed by the comparison of the pre- and post-equalization data.

3.1. Recording configuration and HeadZap measurement procedure

The HeadZap system allows variable grid densities (spatial resolution) as part of its initial setup. For these experiments, measurements were performed every 10° in the two directions in space, azimuth and elevation. The tolerance for the subject’s head movements was $\pm 1^\circ$ in azimuth around the targeted azimuth, as measured with the head tracker device. Visual feedback was provided to the subject about his/her alignment via the “Aimer” (see section 1.5).

The preparation of a measurement session involved qualifying the subject for the procedure, inserting the blocked-meatus Panasonic WM-61 microphones equipped with foam E-A-R ring seals to ensure adequate positioning and fit, and giving the subject instruction regarding his/her role in the measurement process.

The subject was seated on a rotating stool equipped with a customized chin rest adjustable in height to ensure correct alignment of the ears at the level of the reference 0° elevation and control of pitch and roll. After insertion of the microphones, the microphone leads were fixed to the back of a headband using Velcro patches. The sensor of the head tracker was positioned on top of the subject’s head prior to the measurement so that it aligned directly with his/her vertical axis.

The tracker calibration phase was performed with the subject facing the speaker located at 0° azimuth, 0° elevation, the nominal position defined as directly in front of a listener. After checking the subject’s alignment, the position was recorded and served as a reference location for further measurements. A sound test was performed by having the subject rotate so that the right ear was oriented toward the speaker at 0° , 0° (subject facing -90° azimuth) to measure the SNR ratio (SNR should be at least 50 dB, average 85 dB) and to ensure the correct left/right positioning of the microphones.

The measurement phase began with the subject rotating so that he/she faced 180° azimuth. The subject then rotated clockwise from 180° to the next azimuth position at intervals of 10° . The measurements were triggered manually, allowing the firing of the test signals only when the subject’s position was stable. For each azimuth position, measurements for the left and right ears were made in rapid succession at all 12 elevations. This process was repeated until all 36 positions in azimuth were measured. If a problem was encountered during the measurements for a given azimuth position, the measurements could be repeated until a satisfactory result was achieved. The data were stored in an HRTF database and in a raw HRIR response file.

After measuring the HRIRs, headphone response measurements were made using the utility *PhoneQ* (described in Section 1.9). The signal level was adjusted to provide a comfortable level. Headphone responses were measured by having the subject put the headphones on and off three times. The three responses were then averaged to reduce the potential contribution

of anomalies due to a single response, thus providing a more accurate estimate of the headphone response. The headphone equalization measures as well as the free-field measures were then applied to the initial file and the data were normalized. In preparation for rendering via the SLAB system, the initial sampling rate of 96 kHz was down-sampled to 44.1 kHz, the filter taps were reduced from 256 to 128 points, and the final data were saved in the SLAB format.

3.2. HRTF analysis and repeatability

Three subjects, aged 28 to 45, participated in the study. The subject pool included two of the authors (*MA* and *MG*). All had normal hearing with age-related variations. Subjects *MA* and *MG* performed 3 non-consecutive (> one hour interval) measurements that required complete repositioning of the subject and the microphones.

HRTF measurement data verification was performed using the “slabtools” collection of MATLAB utilities, a component of the SLAB User Release [1]. The three primary tools used were hlab() (“HRTF Lab”), cbe() (“Critical Band Energy”), and hcom() (“HRTF Comparison”). hlab() is a MATLAB application that provides a variety of HRTF frequency domain, time domain, and energy visualizations. These visualizations were used to detect trends, expected and erroneous, in the measured and processed data. In Figure 15, hlab() is used to compare two raw blocked-meatus measurements. The measured impulse response length is 1024 samples at 96kHz and begins at the playback of the Golay code. It includes speaker-to-mic propagation delay, the HRIR, and room reflections. The visualization selected is left ear magnitude versus azimuth at 0 elevation. Since the measured data are not yet time-windowed to remove reflections (section 2.1), a constant azimuth-independent comb filter effect can be seen for all azimuths in the dark vertical lines in the plots. Also shown, are the expected azimuth-dependent pinna notches seen as rising or falling patterns in the magnitude data. The detection of erroneous trends revealed several bugs that were later fixed during the shakedown and testing phase of HeadZap development.

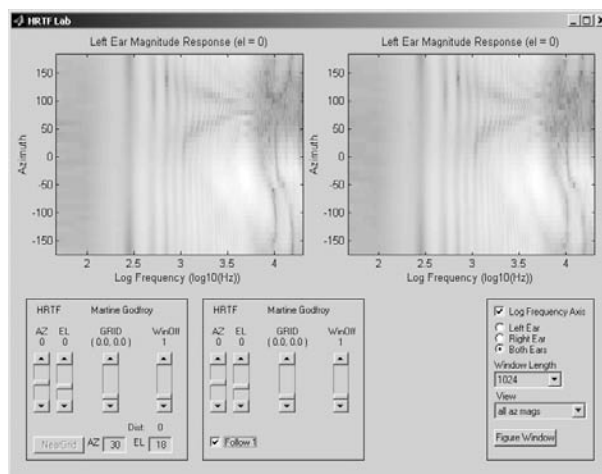


Figure 15. HRTF lab screenshot illustrating trend analysis in HRTF databases.

To simulate the critical band smoothing that determines the frequency resolution of the human auditory system, Slaney’s Auditory Toolbox [11] implementation of the Equivalent Rectangular Bandwidth (ERB) scale was applied to the data. The slabtool cbe() passes the HRTFs through a gammatone filterbank that divides the signal into 20 ERB bands in the audible frequency range and then calculates the HRTF energy in each band. Critical band processing results in a perceptually-based smoothing of frequency domain data. This greatly reduces the number of frequency domain data points and improves the perceptual relevance of HRTF comparisons. When comparing unsmoothed data, high frequency notches that differ by only a few Hertz can appear significant. But in the context of human perception, the difference may be negligible. Figure 16 illustrates the variation in energy level across ERB bands for three different HRTF measurements for subject *MG*. As can be seen from the graph, intra-subject variability is minimal for this specific case and suggests that a high degree of measurement repeatability can be achieved.

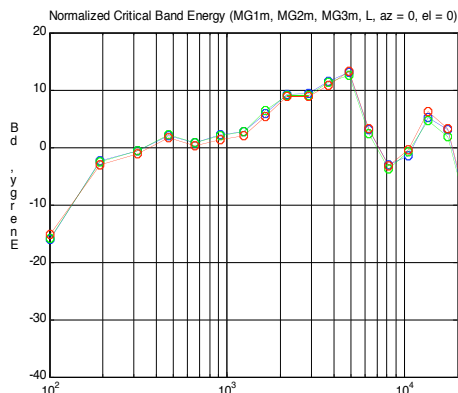


Figure 16. Variations in the critical band energy level (dB) for the measured HRTFs as a function of frequency for subject *MG*, left ear. The plot shows the data for three repetitions of the measurements for a source located at (0°, 0°).

In order to assess repeatability across an entire HRTF map, a visualization was developed to quickly compare differences in Critical Band Energy (CBE) across azimuth and elevation. The `slabtool hcom()` is used to visualize HRTF similarity in a single surface plot. For two given HRTFs, the critical band energy (CBE) log magnitudes are differenced (i.e., a spectral energy ratio). The root-mean-square (RMS) error of the resulting difference sequence is used as a comparison metric for the two HRTFs:

$$\text{RMS Error} = \sqrt{\sum ((\text{CBE1} - \text{CBE2})^2) / (N-1)} \quad (3.1)$$

This process is performed for all HRTFs in a database and plotted by azimuth and elevation on a surface plot. The overall height of this plot indicates the overall difference between the two HRTF databases. Isolated peaks can indicate idiosyncratic differences and, potentially, a bad measurement.

RMS errors were computed for each of the three possible pairs of the repeated HRTFs measurements for two subjects. Data were also segregated into two categories, the measured data (m) that are the raw recorded impulse responses measured at the blocked-meatus microphones and the processed data, (p) the SLAB-format data that includes post-processing described in section 1. This distinction allows separate estimates of the error due to the measurement process alone and the error due to post-processing, such as headphone repositioning. Table I summarizes RMS error metrics for the three HRTF measurements of the

two subjects. Figures 17 and 18 visualize repeatability of the measured and processed data for HRTF comparisons *MG1/MG2* and *MA2/MA3* in Table I.

Comparison	m/p	Mean RMS error	Min	Max
<i>MA1/MA2</i>	m	5.93 (SD=0.6)	4.82	7.94
	p	4.28 (SD=0.90)	2.39	6.9
<i>MA1/MA3</i>	m	5.11 (SD=0.55)	3.8	7.05
	p	3.06 (SD=0.63)	1.87	5.78
<i>MA2/MA3</i>	m	2.46 (SD=0.55)	1.06	4.5
	p	2.83 (SD=0.76)	1.63	5.02
<i>MG1/MG2</i>	m	0.85 (SD=0.48)	0.23	5.55
	p	0.66 (SD=0.25)	0.23	1.73
<i>MG1/MG3</i>	m	1.23 (SD=0.5)	0.55	6.63
	p	2.49 (SD=0.5)	1.58	4.19
<i>MG2/MG3</i>	m	1.32 (SD=0.58)	0.61	7.53
	p	2.37 (SD=0.43)	1.41	3.71

Table I. Mean RMS errors and Standard Deviation for Subjects *MA* and *MG* cross all conditions, measured (m) vs. processed data (p).

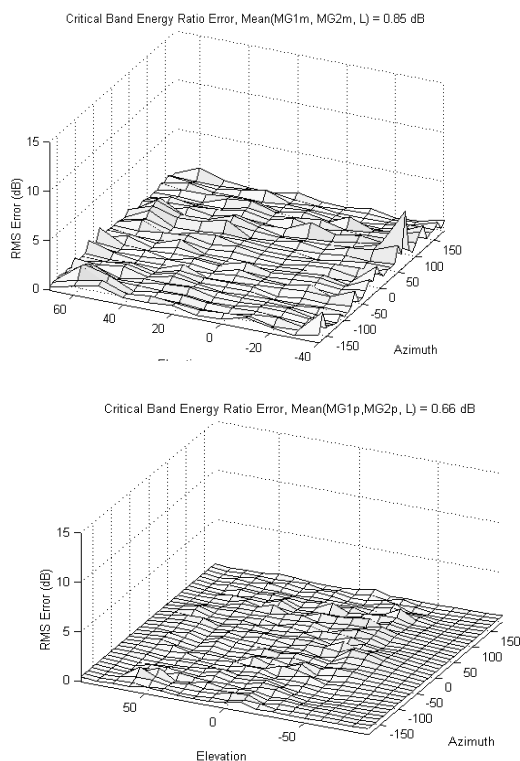


Figure 17. CBE ratio error for two of the HRTF measurements for subject *MG* is shown as a surface plot for all source locations. Top: measured data (.m); bottom: processed data (.p).

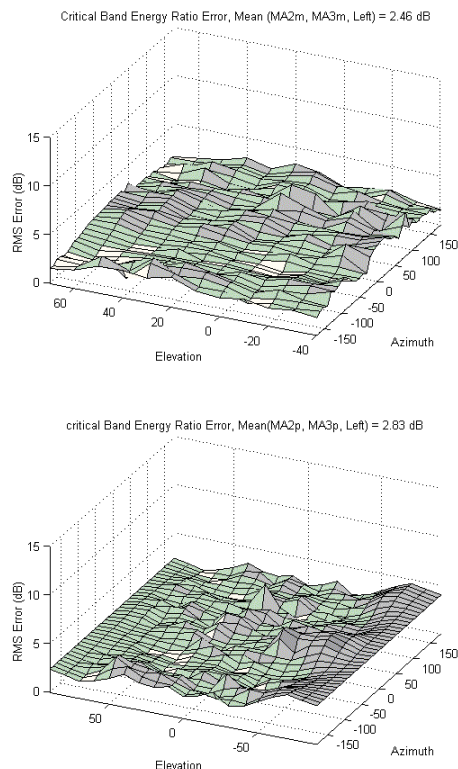


Figure 18. CBE ratio error for two of the HRTF measurements for subject *MA* is shown as a surface plot for all source locations. Top: measured data (.m); bottom: processed data (.p).

As seen in Table I, there is no clear trend for whether the measured or processed measurements are more or less repeatable. Subject *MA*'s repeatability metrics tended to improve with processing while subject *MG*'s metrics showed mixed results between the measured and processed metrics. The lack of a trend in the measured vs. processed data can be explained by three types of processing that may improve or degrade the metric depending on the measurement: time-windowing, bass extension, and headphone equalization.

4. PERCEPTUAL HRTF EVALUATION

In this section, behavioral data measuring sound localization performance for human listeners is presented using HRTFs measured with the HeadZap system. The overall performance is discussed as a function of the direction and the eccentricity of the stimulus in the observer's perceptive field. A

comparison with free-field data is performed using previous data collected from subject *MG*.

4.1. Method

The experiment presented in this paper is a replication of a previous experiment [13] designed to investigate two-dimensional unimodal visual, auditory, and bimodal visual-auditory localization in the free-field, with the primary difference being the type of sound source used to deliver the auditory stimulus. In the original experiment, a loudspeaker was used; in the current experiment, a virtual sound source simulated by the SLAB system and using HRTFs measured with the HeadZap system is used. The experiment is an ongoing evaluation of multimodal interactions in localization. Here, some of the current data are presented to illustrate one type of application of the HeadZap measurement system, and to exemplify the veridicality of the measured spatial auditory cues.

Subjects were positioned in the center of a 125 cm diameter spherical dome, with their head position maintained level by a chinrest (see Figure 19). They wore headphones that provided a virtual auditory stimulus using either their own HRTFs (individualized condition) or random, non-individualized HRTFs (from other participants, also measured with the HeadZap system). The stimuli (targets) were 100 ms in duration and consisted of a broadband noise burst, a spot of light, or the combination of the two. Targets were presented randomly in 29 locations between $\pm 15^\circ$ azimuth and elevation; the visual stimuli were projected on the interior surface of the dome.

For each trial, after presentation of the target, a visual cursor (a +) projected on the dome surface was moved by the subject toward its perceived position, via a track ball mouse positioned on his/her leg. This enabled a motor-decoupled measurement of the spatial percept (as opposed to using, e.g., a pointing response). Data were gathered from 290 trials per subject (a random presentation of 10 repetitions for each target position and condition) under open-loop conditions (no acoustic, visual or verbal feedback was provided). The localization performance was analyzed for three subjects for the auditory modality condition only (the ongoing study involves combinations of visual and auditory stimuli). Also, a comparative analysis is presented for subject *MG* who participated in the original 2004 study [13]. This provides a direct comparison of localization of actual sound sources to

virtual sound localization performance, using the spatial auditory cues measured by HeadZap.

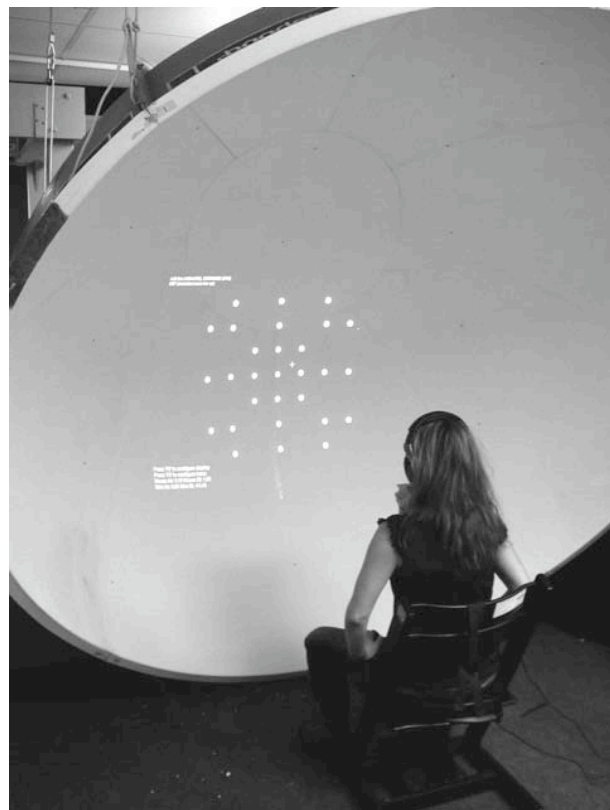


Figure 19. Subject with projected visual targets within the dome. Here, all the possible locations for the visual and auditory stimuli are shown. A visual cursor (+) is moved to the perceived memorized location of the target.

4.2. Data Analysis

Individual regression analyses were performed for each subject to describe the general trend of their responses and to give an estimate of the optimal linear fit between the targets and the responses overall. *Precision* is defined here as the combination of the correlation coefficient (or gain) from these regression analyses, and by the *constant error* in localization computed as the difference (in degrees) between the average final position of the cursor relative to the target in azimuth and in elevation. *Constant error* is defined by the angular distance and orientation relative to the target, independent of direction (see Figure 20, top row).

The *accuracy* (bias) is expressed as the mean signed error value, described here as ‘overshoot’ (positive

values associated with errors eccentric to the target with respect to the reference coordinates) or ‘undershoot’ (negative values associated with errors central to the target). The accuracy of the responses is plotted in Figure 20 (middle row) by connecting the mean responses for each location. From these plots, one can also see the degree of spatial warping or compression that occurs. *Dispersion* refers to an estimation of the maximum variance (σ_b) and minimum variance (σ_a) of the data distribution along the two main Cartesian dimensions, azimuth and elevation. It has been shown previously [14] that a decomposition of responses along azimuth and elevation axes fails to take into account the main orientation of the data distribution, resulting in an erroneous representation of the bi-dimensional response patterns. For this study, the minimum and maximum Eigenvalues (oriented dispersion) were computed to provide the variance of the distributions. Ellipses were drawn using the maximum variance as the Eigenvector (main orientation of the distribution). See Figure 20 (bottom row).

4.3. Results

4.3.1. Overall Angular Error

The values of the correlation coefficients differ as a function of azimuth and elevation, as seen in Table II. The subject's mean judgments showed very high correlation (0.87 to 0.97) with the actual azimuthal location of the sound. In elevation, the correlation is lower ($< .6$). The angular error (Figure 20, top) is dependent on direction, with the largest errors for targets on the vertical axis at 0° azimuth (mean 7.55°) or for those that are not on the azimuthal plane (7.26°).

For targets on the azimuthal plane, the angular error drops drastically to a mean value of 2.9° . Angular error varies only slightly as a function of eccentricity in azimuth, and only the eccentricities $\pm 5^\circ$ are significantly different (Scheffe: $p=0.0355$). In elevation, the angular error increases with the degree of eccentricity of the sound source. For this case there is no difference in error as a function of the quality of the HRTFs, Generic vs. individualized ($F_{1,108}=1.109$; $p=.7422$). This may have occurred as a result of the random presentation of the type of HRTF used (individualized versus generic), preventing the subject from “learning” the properties of the spatial cue. Consequently, Figure 20 and Table II summarize the data for the virtual sound sources with the generic and individualized HRTF stimuli combined.

4.3.4. Variance of the responses

		BUM	MG	MA
	Direction	Mean (SD)	Mean (S.D)	Mean(SD)
Gain	Azimuth	.87	.97	.92
	Elevation	.25	.53	.59
Bias [Deg]	Azimuth	-.19 (3.34)	.34 (1.84)	.14 (2.76)
	Elevation	3.29 (8.26)	.12 (6.67)	.63 (6.25)
Constant Error [Deg]	Azimuth	2.52 (2.19)	1.34 (1.3)	1.97 (1.93)
	Elevation	7.01 (5.33)	4.81 (4.7)	4.42 (4.46)
Eig.Min		5.6 (5.8)	4.04 (4.32)	9.09 (10.09)
Eig.Max		23.2 (17.39)	21.7 (14.8)	33.6 (22.79)

Table II. *Precision* (gain), *accuracy* (bias) and *dispersion* (constant error and Eigenvalues) of the localization responses for the 3 subjects as a function of the two main directions, azimuth vs. elevation.

4.3.2. Azimuth Localization performance

Analysis of the overall azimuth localization performance shown in Table II indicates inter-individual variations in the signed azimuth error (repeated measures MANOVA: $F_{2, 450}=4.525$; $p=0.0113$; Scheffe test: $BUM, MG=0.0317^\circ$; $BUM, MA=0.25^\circ$; $MA, MG=0.61^\circ$), an effect of eccentricity ($F_{6, 450}=118.459$; $p<0.0001$), as well as the interaction effect between azimuth and eccentricity ($F_{12, 450}=8.29$; $p<0.0001$). The variation between subjects is characterized primarily by the direction of the bias. Indeed, subject *BUM* presents a tendency to underestimate the position of the sound. Conversely, subjects *MA* and *MG* show a slight overshoot of the target location.

4.3.3. Elevation localization performance

Accuracy in elevation follows a somewhat different evolution since the global analysis underlines a significant difference in amount of bias between the 3 subjects, as attested by the differences in warping (compression) of the auditory space in Figure 20 (paired MANOVA: $F_{2, 450}=154.75$; $p<0.0001$, Scheffe test: all the differences significant at least at $p=0.0031$). In this dimension, the amount of bias increases with the amount of eccentricity ($F_{6, 450}=738.495$; $p<0.0001$). Despite this systematic undershooting (negative values associated with errors central to the target) and variations in sensitivity (Eccentricity*Accuracy, $F_{12, 450}=13.38$; $p<0.0001$), subjects nevertheless perceived the differences in the elevation between the sources.

The dispersion of the responses (constant error) is significantly lower in azimuth than in elevation, which reaches almost 3 times the azimuth value (see Table II). To obtain a two-dimensional map of the localization variability, the minimum and the maximum variance (Eigenvalues) in the pointing errors were visualized as a polar plot oriented along the maximum variance, also called the Eigenvector. The resultant can be described as ellipses, with various degrees of anisotropy and orientations. A differential effect of elevation on the minimum and maximum variance can be observed; only the minimum variance decreases with elevation. Increasing eccentricity in azimuth, conversely, reduces the variance in both dimensions and the smallest variance is observed for the most peripheral targets.

4.4. Intra individual analysis of localization performance for real and virtual sound sources

The design of the experiment presented in section 4.1 also allowed a paired comparison (for subject *MG*) of free-field (real loudspeaker) and virtual auditory (HRTF-filtered simulation of the real loudspeaker) conditions for 9 auditory targets positions in the central frontal hemi-field (azimuth $-10^\circ, 0^\circ, +10^\circ$; elevations $-10^\circ, 0^\circ, +10^\circ$). Figure 21 indicates the results in terms of ellipses (variance), vectors (precision) and stars (accuracy). The following analyses combine localization data for generic and personalized HRTFs.

4.4.1. Precision

The overall angular error is directly dependent on the location of the target, in terms of azimuth and elevation. Between free-field auditory (FFA) and virtual auditory (VA) conditions, the responses as expressed by the correlation coefficient remains very similar in azimuth (FFA: $R^2=.99$, residual error $=0.42^\circ$; VA: $R^2=.98$, residual error $=1.14^\circ$). For elevation, the decrease of the correlation coefficient in the virtual auditory condition (FFA: $R^2=.64$; residual error $=4.04^\circ$; VA, $R^2=.56$; residual error $=1.4^\circ$) indicates a global underestimation of the sound position in the virtual condition, compared to the free-field condition. It should be noted that despite a weaker correlation coefficient for elevation in the VA condition, the residual error is more than 50% smaller than in the FFA condition, expressing more homogeneous response patterns.

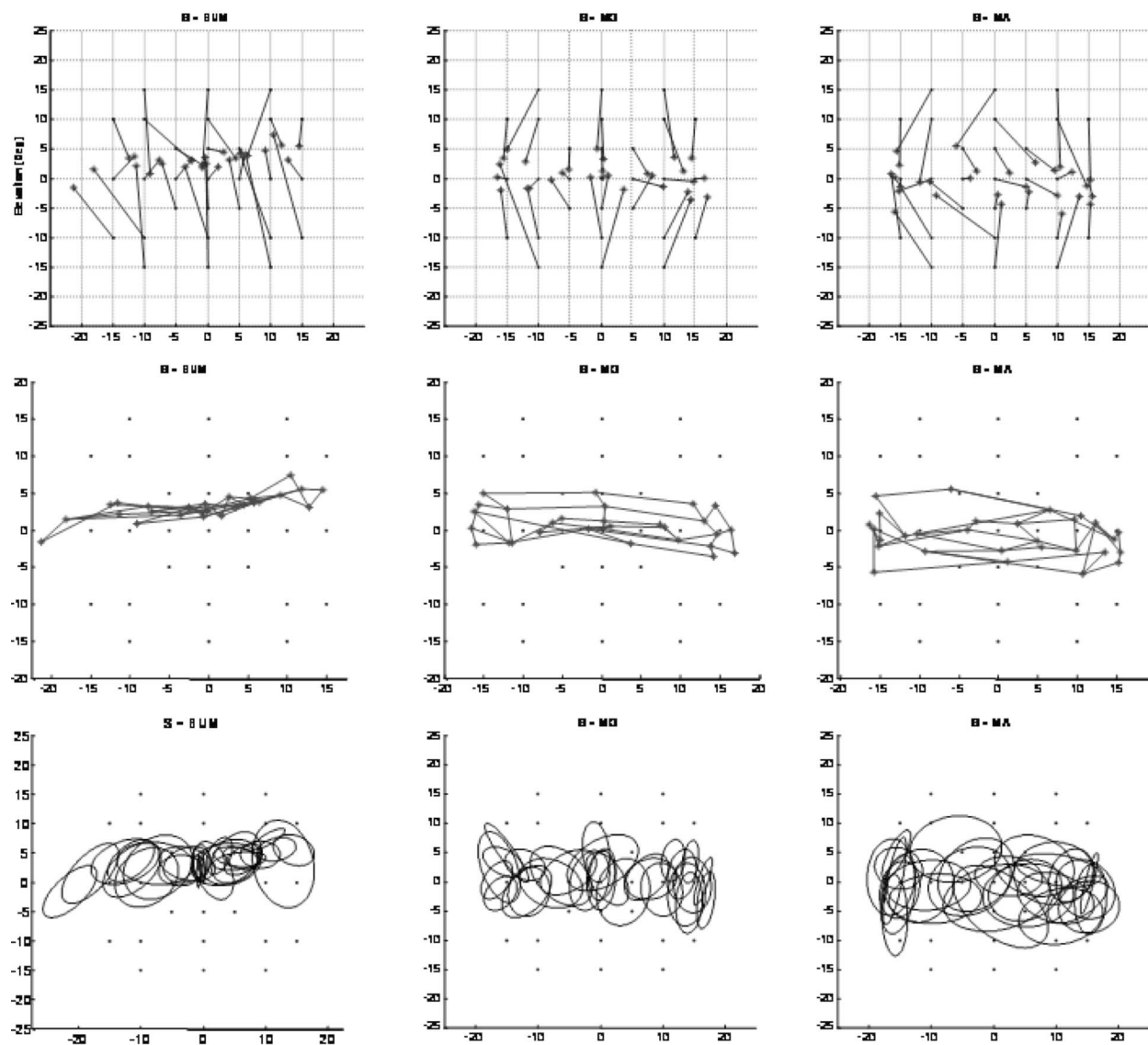


Figure 20. Amplitude of oriented error (top row), perceptual space compression (middle row) and variance of the responses (Eigenvalues, bottom row) for the 3 subjects *BUM*, *MG* and *MA* from left to right. The x-axis is the response azimuth and the y-axis is the response elevation in degrees.

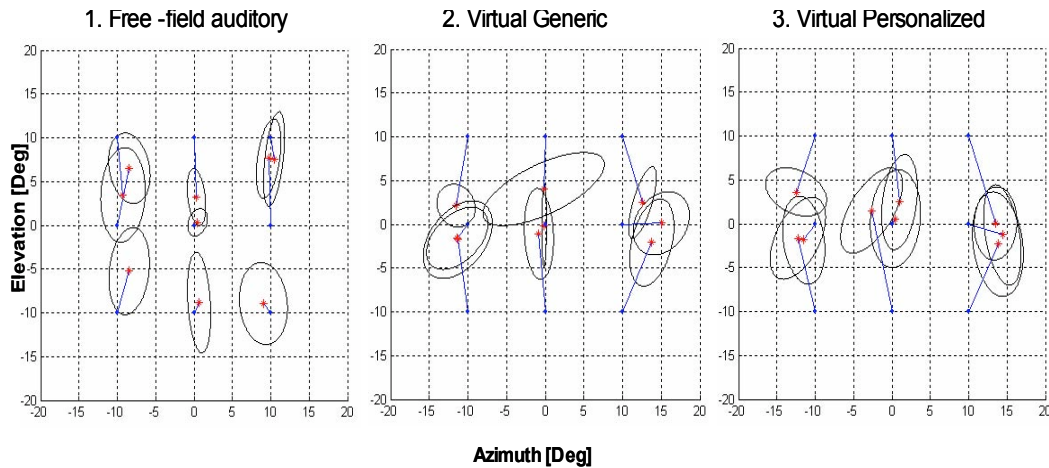


Figure 21. Response patterns for real sound localization (left) and virtual sound localization (middle: generic HRTFs and right: personalized HRTFs) of subject *MG*.

4.4.2.Accuracy

In azimuth, the interaction between conditions (FFA versus VA) and eccentricity (-10° , 0° , $+10^\circ$) is significant ($F_{2,15}=65.92$; $p<0.0001$); see Table III. This is related to the fact that (1) the accuracy is not symmetrical around the vertical (median sagittal) plane, especially in the VA condition where the error in the right hemi-field is twice the error in the contra lateral (left) hemi-field; and (2) the localization of the azimuth was underestimated in the FFA condition while overestimated in the VA condition.

In elevation, the interaction between condition (FFA versus VA) and eccentricity (-10° , 0° , $+10^\circ$) is also significant ($F_{2,15}=19.39$; $p<0.0001$) and is related to a significant improvement in accuracy in the virtual auditory condition for the localization of targets at 0° elevation, as shown in Table III. The perception of a subjective “auditory horizon” (perception of the horizontal median plane) is more precise and more stable (reduced variance) than in the free-field sound condition. Note also the symmetry in response accuracy around the vertical (median sagittal) plane in the VA condition.

4.4.3.Dispersion

The maximum variance (Eigenvalue max) in the localization performance is not significantly different

between the FFA (mean= 21.21° , sd= 9.48°) and the VA conditions (mean= 27.13° , sd= 13.12°). The maximum Eigenvalue increases slightly (but is still insignificant) as a function of the eccentricity in elevation, but remains constant whatever the eccentricity in azimuth. Conversely, the minimum variance (Eigenvalue minimum) significantly differs between the real and the virtual presentation ($F_{1,17}=6.91$; $p=.01$), mean RS=4.4, sd=3.28; mean VS=6.74, sd=3.05, and as a function of the eccentricity in azimuth only (no effect of the eccentricity in elevation).

	Azimuth		P (Scheffe test)	Elevation		P (Scheffe test)
	FFA	VA		FFA	VA	
Eccentricity						
-10°	1.27 (.42)	-1.75 (4.99)	<.0001	2.33 (1.88)	8.76 (1.41)	.0018
0°	0.47 (0.15)	-0.4 (1.26)	.18	4.82 (3.51)	-0.69 (1.01)	.0238
$+10^\circ$	-0.27 (.62)	3.84 (.82)	.0004	-5.23 (2.19)	-7.53 (1.4)	.1487

Table III. Accuracy of localization (degrees) as a function of eccentricity and direction (azimuth vs. elevation) for free-field acoustic (FFA) and virtual acoustic (VA) conditions.

This difference is related to two factors: (1) the left/right asymmetry of the variance in the real sound condition (effect of eccentricity in azimuth for RS, $F_{2,15}=8.26$; $p=.003$); and (2) the loss of the singularity in the virtual sound condition for the target located at 0° azimuth (no effect of eccentricity in azimuth for VA, $F_{2,15}=0.44$; $p=.64$), as expressed in Figure 21 (left) by the maximum anisotropy (Eigenvalue min/Eigenvalue max) of the ellipses in the real sound condition.

4.5. Discussion

Within the limited range of sound source positions presented here, the data indicate that the localization performance of virtual sources presented under headphones for inexperienced subjects provided a very accurate perception of azimuth. In elevation, we observed a compression of the auditory space around the horizontal median plane, previously observed to a lesser extent in the free-field source condition. The evidence of differentiation among very proximal elevations (5°) is an argument for successful stream segregation.

5. CONCLUSION

This paper has approached the verification of the HeadZap system for measuring HRTFs from both a system design and psychoacoustic applications standpoint. We are currently involved in further exploration that includes comparative measurements between different laboratories around the world, using a single dummy head microphone system (a project dubbed 'Club Fritz' in honor of the Neumann KU-100).

6. ACKNOWLEDGEMENTS

This work was supported by NASA's Airspace Systems Program and the Space Human Factors Engineering Program. We would like to thank our colleagues and the lead developer of the HeadZap system, William Chapin, as well as the developer of the original SNAPSHOT system, Jonathan Abel, whose work influenced many aspects of the current HeadZap system.

7. REFERENCES

- [1] E. M. Wenzel, M. Arruda, D. J. Kistler and F. L. Wightman, "Localization using non-individualized head-related transfer functions," *J. Acoust. Soc. Am.*, vol. 94, pp. 111-123 (1993).
- [2] E. M. Wenzel, J. D. Miller and J. S. Abel, "Sound Lab: A real-time, software-based system for the study of spatial hearing," *Proc. 108th Conv. Aud. Eng. Soc.*, Paris, Feb. 2000, New York: Audio Engineering Society, Preprint 5140.
- [3] J. D. Miller and E. M. Wenzel, "Recent developments in SLAB: A software-based system for interactive spatial sound synthesis," *ICAD 2002, Int. Conf. Aud. Displ.*, Kyoto, JAPAN, July 2-5 (2002).
- [4] <http://humanfactors.arc.nasa.gov/SLAB>
- [5] R. G. Klump and H. R. Eady, "Some measurements of interaural time difference thresholds," *J. Acoust. Soc. Am.*, vol. 28, pp. 859-860 (1956).
- [6] R. H. Dye and E. R. Hafter, "The effects of intensity on the detection of interaural differences of time in high-frequency trains of clicks," *J. Acoust. Soc. Am.*, vol. 75, pp. 1593-1598 (1984).
- [7] J. V. Tobias and E. D. Schubert, "Effective onset duration of auditory stimuli," *J. Acoust. Soc. Am.*, vol. 31, pp. 1595-1605 (1954).
- [8] B. C. J. Moore, *An Introduction to the Psychology of Hearing*. Elsevier, Ltd: London (1994).
- [9] M. J. E. Golay, "Complementary series," *IRE Trans. Info. Th.*, vol. 7, pp. 82-87 (1961).
- [10] D. T. Sandwell, "Bi-harmonic spline interpolation of GEOS-3 and SEASAT altimeter data," *Geophysical Research Letters*, vol. 2, pp. 139-142 (1987).
- [11] F. L. Wightman and D. J. Kistler, "Headphone simulation of free-field listening. I: Stimulus synthesis," *J. Acoust. Soc. Am.*, vol. 85, pp. 858-67 (1989).
- [12] M. Slaney, *Auditory Toolbox* <http://rvl4.ecn.purdue.edu/~malcolm/interval/1998-010/>
- [13] M. Godfroy and C. Roumes, "Multisensory enhancement of localization with synergistic visual-auditory cues," *Proc. 26th COGSCI*, Chicago, IL (2004).
- [14] M. Godfroy and R. B. Welch, "Two-dimensional localization with spatially and temporally congruent visual-auditory cues," *Proc 46th Conf. Psychon. Soc.*, Toronto, pp. 10-13 (Nov., 2005).