



A Virtual Audio Environment for Testing Dummy-Head HRTFs modeling Real Life Situations

György Wersényi
Széchenyi István University, Hungary.

József Répás
Széchenyi István University, Hungary.

Summary

Virtual audio simulators usually incorporate HRTF filtering and headphone playback. The most important parameters for simulation include accuracy and spatial resolution of the applied HRTFs, setting the individual parameters (customization) and further signals processing algorithms in order to equalize the headphone or tracking head movements. This paper presents a custom built MATLAB-based virtual audio environment for listening tests using various dummy-head HRTFs, ITD setting methods, headphone equalization etc. Furthermore, first results from a listening test for comparison of HRTFs recorded with a manikin wearing hair or glasses are also presented.

PACS no. 43.66Qp, 43.66Pn

1. Introduction

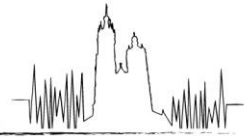
The use of HRTFs in virtual audio has been an investigated field for a long time [1]-[3]. This mainly focuses on the measurement method and data collection. Spatial resolution, measurement accuracy and repeatability, signal-to-ratio issues, individuality are the most important questions [4]-[5]. Furthermore, representation and data formats, scaling methods, filter realizations also play a significant role during playback [6]-[8]. Simulators may also include different methods for customization, such as settings of anthropometric measures (head or pinna size), selection methods for the “best fitting” HRTF set, headphone equalization or even tracking head movements. Listening tests aim to test the localization performance and errors, in-the-head localization or front-back-reversal rates and subjective evaluation. Especially in the early 90s this area offered lot of research work and results of the binaural technique indicated parameters such as individually measured HRTFs, good resolution and accuracy in frequency and space to be very important [3], [8], [9]. That is, the generally decreased localization performance in virtual audio was suggested to be – among others - due to inaccurately measured

HRTFs and differences compared to individual HRTFs.

Beside HRTF filtering, other important parameter for the simulation is the time difference between the two earsignals in case of a sound source outside of the median plane: the interaural time difference (ITD). It is a usual method to assume the HRTF to be a minimalphase filter, that is, a realization of a filter corresponding to the magnitude response and a pure time delay during playback can result in sufficient localization.

Our former research tested whether differences and disturbances near the head have significant influence on the fine structure of the HRTFs [10], [11]. An accurate dummy-head measurement system was introduced and a huge database of HRTFs was recorded using the manikin equipped with hair, glasses, caps, clothing etc. The objective evaluation revealed significant effect of these in given directions and frequency ranges: differences up to 20 dB could be detected from the same sound source direction in comparison of the naked and dressed torso’s HRTF. The other question if this is audible in any ways and what kind of influence this has during virtual localization has not been tested, mostly due to the missing simulator program back then.

In order to test localization performance in listening tests and to be able to set various environmental conditions a custom made simulator was programmed and has been continuously



updated on the MATLAB platform [12]. After testing its functionality and debugging a series of experiments have been designed to look deeper into the audibility of artifacts using different HRTF sets. This paper presents briefly the functionality of the virtual audio environment including the GUI, settings of the HRTFs and ITD information based on head diameter and different approximations, headphone equalization and even simulating distance information of reflecting surfaces. Furthermore, latest results from the first comparative listening test using HRTFs with hair, glasses and baseball cap are also presented.

2. Measurement setup

2.1 The virtual audio environment

The virtual audio simulator, formerly referred as the VAS, was developed in the MATLAB programming environment. The HRTF dataset was recorded earlier using the Brüel&Kjaer 4128C dummy-head with built in microphones at the eardrums. High spatial resolution (1 degree horizontally and 5 degrees vertically in some regions) and high signal-to-noise ratio was achieved that resulted in high measurement accuracy and repeatability (about 1 dB) [10], [11]. A huge database of HRTFs was recorded in different environmental conditions. In these conditions, HRTFs of the dummy wearing hair, cap, clothing, glasses were also measured and compared. Because of the unique data format of the measured HRTFs a dedicated playback system had to be developed for the simulation platform.

Figure 1 shows the screenshot of the current status of the GUI that is used for the listening tests. Mono or stereo wave files can be loaded into the system and played back once or looped. The time function and spectrum can be displayed for control purposes. As default, 13 cm for head diameter is set, but it can be adjusted individually. Similarly, the default estimation for ITD information will be calculated using the Woodworth formula [8], [13], [14]. However, for further experiments, the estimation method of Kuhn can be applied as well [15], [16]. On the right side, the direction of the sound source can be set in one degree spatial accuracy in the horizontal plane as a single steady source or as a moving sound source around the head. The applied HRTFs for the filtering are also displayed. The filtering is realized in the frequency domain by multiplication of the amplitude response only, followed by the appropriate ITD-

delay between the two ears in the time domain. The resulting stereo wave file can be played back or saved. Although currently not used, reflections and elevation can be also added to the simulation easily.

2.2. Settings of ITD

Setting of ITD information is an important stage during simulation of sound source directions. The software has the following different possibilities implemented for ITD estimation. The default setting is the Woodworth-formula:

$$ITD = \frac{d(\varphi + \sin \varphi) \cos \delta}{2c} \quad (1)$$

This formula can be used in the entire frequency range both for elevation and azimuth. The software also allows using the Kuhn-formula:

$$ITD_{lowf} \approx \frac{3a}{c} \sin \varphi \quad (2)$$

$$ITD_{highf} \approx \frac{2a}{c} \sin \varphi \quad (3)$$

If the frequency range is below 500 Hz, the low frequency formula can be used. For frequencies above 2000 Hz, the high frequency formula can be applied. Between these values, the ITD is frequency dependent with a slight decreasing profile. Kuhn, however, also developed a formula that is independent of frequency and contains also elevation information:

$$ITD \approx \frac{a\varphi}{c} (\arcsin(\cos \delta \sin \varphi) + \cos \delta \sin \varphi) \quad (4)$$

All of these formulas estimate the ITD based on a rigid sphere head model, where d is diameter, a is the radius, c is speed of sound, φ is azimuth, δ is elevation in degrees.

The playback environment does not include the headphone equalization module directly. The applied Sennheiser HD650 headphone was measured using the same dummy-head. Its frequency response for both sides were measured ten times, averaged and equalized by an inverse FIR-filter in MATLAB prior to the listening tests [12]. The excitation signal meant for the listening test (in this case a 5 sec white noise sample) was pre-filtered with the equalization filter and can be used as excitation directly loaded into the system.

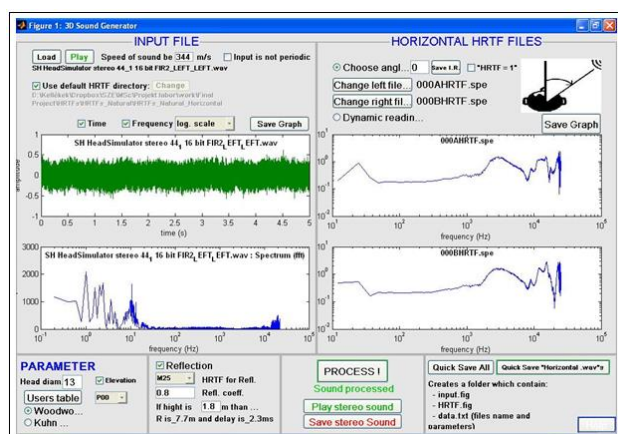
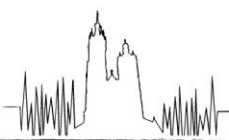


Figure 1. Screenshot of the actual version of the simulator program.

2.3. Setup of the listening tests

The listening test was installed in the anechoic chamber of the university. The first session included a test with the following parameters and restrictions:

- 5 sec of pre-filtered white noise excitation (resulting in headphone equalization for left and right side respectively),
- measurement of individual head size by measuring the distance between the ear canal entrances on the back side of the skull,
- setting the ITD information based on the Woodworth formula,
- settings of possible sound source directions in the horizontal plane in 10 degrees pacing and for directions -15,0,15,30 and 45° in the median plane.

21 male and 9 female subjects between 10 and 62 years participated (mean 29). Subjects were sitting on a comfortable chair during the session of an absolute localization task. During accommodation time, a detailed description of the procedure was given. Subjects were instructed to call perceived sound direction (10-degree pacing) from the left and right side, however, actual simulated source directions were limited to 16 (Fig. 2). Furthermore, front and back directions were simulated three times in order to determine front-back confusion rates. Conditions included HRTFs from the naked torso, HRTFs recorded with hair, with glasses and with a baseball cap. Source directions were simulated in randomized order. Error rates were collected as deviations from the simulated source direction in degrees as well as in-the-head localization and front-back error rates. As front-back errors are frequent in virtual audio simulations, evaluation of these remains sometimes unnoted. That means,

reported directions symmetrical to the frontal plane (e.g. +10 degrees and +170 degrees) may not be discriminated.

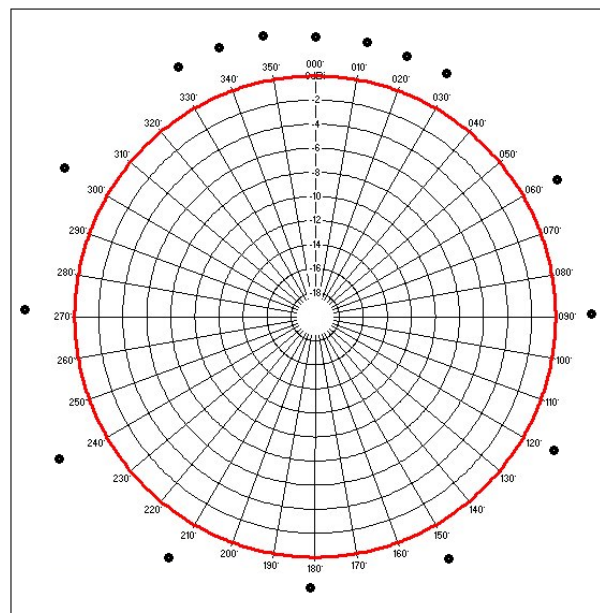


Figure 2. Scheme of the source directions during presentation in the horizontal plane. Dots correspond to actual possible source directions (only 16) unknown to listeners who can report all 36 directions.

3. Results

3.1. Normal HRTFs

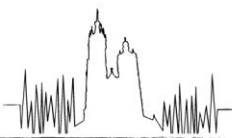
In the horizontal plane 90% of the answers could be evaluated because in 10% of the simulation subjects were not able to determine the direction. From the given answers 30% were correct. The best identification was for direction 270° (63% correct identification) as long the worst identification was for direction 20° (0%).

In case of a frontal source 77% of the answers could be evaluated and in case of a rear source it was 80%. In about 21% of the cases subjects reported in-the-head localization, elevation shift or confused perception.

Front direction was detected only by 19% correctly. 42% of the answers indicated « back » and the rest any other possible directions.

Rear direction was detected by 57% correctly. 15% of the answers indicated « front ».

In the median plane only 40% of the answers could be evaluated. From the given answers 29% were correct. The best identification was for direction 0° (57% correct identification) as long the worst identification was for direction -15° (14%).



3.2. HRTFs with hair

In the horizontal plane 91% of the answers could be evaluated. From the given answers 36% were correct. The best identification was for direction 120° (67% correct identification) as long the worst identification was for directions 10° and 30° (0%).

In case of a frontal source 78% of the answers could be evaluated and in case of a rear source it was 79%. In about 21% of the cases subjects reported in-the-head localization, elevation shift or confused perception.

Front direction was detected only by 30% correctly. 41% of the answers indicated « back » and the rest any other possible directions.

Rear direction was detected by 54% correctly. 2% of the answers indicated « front ».

In the median plane only 44% of the answers could be evaluated. From the given answers 37% were correct. The best identification was for direction 0° (53% correct identification) as long the worst identification was for direction 45° (30%).

3.3. HRTFs with cap

In the horizontal plane 90% of the answers could be evaluated. From the given answers 32% were correct. The best identification was for direction 120° (55% correct identification) as long the worst identification was for direction 20° (0%).

In case of a frontal source 84% of the answers could be evaluated and in case of a rear source it was 78%. In about 19% of the cases subjects reported in-the-head localization, elevation shift or confused perception.

Front direction was detected only by 24% correctly. 59% of the answers indicated « back » and the rest any other possible directions.

Rear direction was detected by 60% correctly. 24% of the answers indicated « front ».

In the median plane only 31% of the answers could be evaluated. From the given answers 42% were correct. The best identification was for direction -15° (57% correct identification) as long the worst identification was for direction 0° (31%).

3.4. HRTFs with glasses

In the horizontal plane 90% of the answers could be evaluated. From the given answers 34% were correct. The best identification was for direction 150° (66% correct identification) as long the worst identification was for direction 30° (3%).

In case of a frontal source 77% of the answers could

be evaluated and in case of a rear source it was 78%. In about 23% of the cases subjects reported in-the-head localization, elevation shift or confused perception.

Front direction was detected only by 26% correctly. 49% of the answers indicated « back » and the rest any other possible directions.

Rear direction was detected by 43% correctly. 12% of the answers indicated « front ».

In the median plane only 37% of the answers could be evaluated. From the given answers 37% were correct. The best identification was for direction 0° (57% correct identification) as long the worst identification was for direction 15° (17%).

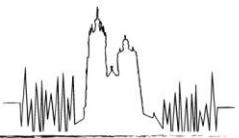
4. Discussion

Results in the horizontal plane show no significant difference among different HRTF sets. Using any of the HRTF sets about 90% of the answers could be used for evaluation. From this, 30-36% were actually correct, that is, only about 27-32% of the answers were correct in the horizontal plane. The most remarkable thing is that source directions near the front between +30° and -30° were the hardest to localize correctly. Furthermore, directions around the sides were detected more easily. A detailed statistical analysis will be needed to test variances, standard deviations of the mean error rates.

In all cases almost 20% reported in-the-head localization, elevation shift or delivered no answer at all. This rate is surprisingly good.

Front and back directions were detected correctly in 19-30% and in 43-60% respectively indicating large front-back confusion rates, as expected. Furthermore, there is more error in case of a frontal source.

In the median plane, decreased localization performance was measured with only 37-44% of evaluable answers, that is, only about 10-14% were actually correct. Front direction (0° elevation) was identified the best excluding the case HRTF with cap. It was suggested that shadowing effects caused by the head or any other object near the head may influence median plane localization. Although objective measurements supported that baseball caps do influence HRTFs from selected directions and shadowing effect of the visor could be detected, no detectable difference in localization appeared neither in the median plane nor in the horizontal plane. Nevertheless, source directions outside these two planes and/or at higher elevations than +45° may lead to more localization errors.



Absolute localization errors, in-the-head localization rates and front-back errors are almost independent of the applied HRTF set and quite large. The same can be observed for the median plane where – supporting our former results – vertical localization can be a total failure.

Generally, subjects could not hear any better or worse using HRTFs recorded on the naked manikin or with hair, glasses or a cap. The previously reported differences in the fine structure of the HRTFs caused by these conditions can be detected by the measurement system and analysis, but their influence on localization is not reflected in audible effects or artifacts.

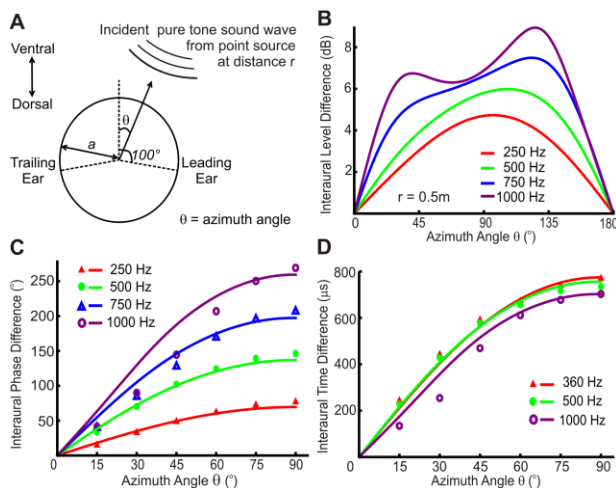


Figure 3. Model predictions for human ILDs and ITDs [17]. A, model to determine ITD or ILD variation with azimuth angle θ for the experimental set up in Mills, Schmidt *et al.*, and Kuhn. The human head is modelled as a solid sphere. Ears are positioned 100° away from the midline. B, azimuthal variation of interaural level difference (ILD) for sound source at 0.5 m, 250 Hz, 500 Hz, 750 Hz or 1000 Hz, as predicted by our acoustic model for the experimental set up by Mills. C, comparison of predicted curves for interaural phase differences (IPDs) and empirical data points from Mills, $r = 0.5$ m. D, comparison of model interaural time differences (ITDs) with empirical data from Kuhn, $r = 3.0$ m.

Although changes in the environment near the head can affect the fine structure of the HRTFs, differences even up to 10-20 dB remain undetected during listening tests. With other words, spectacled people would not increase their localization performance by using HRTFs recorded on a manikin wearing glasses in virtual simulation. Similarly, long-haired or short-haired persons do not benefit from using the appropriately recorded HRTFs. The main problem could be here that the

use of dummy-head HRTFs already introduces increased localization errors, and by modifying them further will not result in any significant difference. This suggests on one side that HRTFs do not have to be recorded very precisely (resolution in frequency) and on the other side, individual recordings or head tracking to be more influential parameters. Although we did not include individual HRTFs (with and without glasses or hair), it is expected that even in this case, changes in the HRTFs would remain undetected during listening tests.

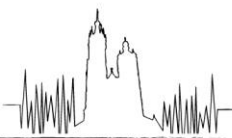
For ITD estimations the Woodworth formula was used. Using other integrated formulas is put to future work but it is assumed this may cause differences in localization. Figure 3 shows recent comparative results of different ITD estimations [17].

5. Conclusions

A MATLAB-based virtual audio simulator was presented suitable for listening tests emulating different environmental conditions mainly by changing the applied HRTF set. The most important goal was to be able to test the previously measured dummy-head HRTF database including HRTFs from the naked and dressed torso for audible effects and artifacts. The first listening session included 30 participants using an equalized headphone, white noise excitation, and simulated sound source directions in the horizontal and vertical plane. HRTFs from the naked torso and HRTFs with glasses, cap and hair were applied. Results indicated that the localization performance of subjects is not sensitive to the fine structure deviations of dummy-head HRTFs caused by these environmental effects. Generally, localization errors were quite large in all situations. Future works includes detailed statistical analysis, testing additional effects of environmental influence (such as reflections simulated via HRTFs) and the role of different estimation methods in the signal processing (ITD formulas, filtering methods of headphone equalization).

Acknowledgement

This research was realized in the frames of TÁMOP 4.2.4. A/2-11-1-2012-0001 „National Excellence Program – Elaborating and operating an inland student and researcher personal support system” The project was subsidized by the European Union and co-financed by the European Social Fund.



References

- [1] J. Blauert: Spatial Hearing. The MIT Press, MA, 1983.
- [2] C. I. Cheng, G. H. Wakefield: Introduction to Head-Related Transfer Functions (HRTFs): Representations of HRTFs in Time, Frequency, and Space. J. Audio Eng. Soc., vol. 49 (2001) 231-249.
- [3] H. Møller, M. F. Sorensen, D. Hammershøi, C. B. Jensen: Head-Related Transfer Functions of human subjects. J. Audio Eng. Soc., vol. 43 (1995) 300-321.
- [4] F. Wightman, D. Kistler: Measurement and validation of human HRTFs for use in hearing research. Acta acustica united with Acustica, vol. 91 (2005) 429-439.
- [5] D. R. Begault, E. Wenzel, M. Anderson: Direct Comparison of the Impact of Head Tracking Reverberation, and Individualized Head-Related Transfer Functions on the Spatial Perception of a Virtual Speech Source. J. Audio Eng. Soc., vol. 49 (2001) 904-917.
- [6] E. M. Wenzel: Localization in virtual acoustic displays. Presence, vol. 1 (1991) 80-107.
- [7] E. M. Wenzel, M. Arruda, D. J. Kistler, F. L. Wightman: Localization using nonindividualized head-related transfer functions. J. Acoust. Soc. Am., vol. 94 (1993) 111-123.
- [8] H. Møller, M. F. Sorensen, C. B. Jensen, D. Hammershøi: Binaural Technique: Do We Need Individual Recordings? J. Audio Eng. Soc., vol. 44 (1996) 451-469.
- [9] H. Møller: "Fundamentals of binaural technology. Applied Acoustics, vol. 36 (1992) 171-218.
- [10] Gy. Wersényi, A. Illényi: Differences in Dummy-Head HRTFs Caused by the Acoustical Environment Near the Head. Electronic Journal of "Technical Acoustics" (EJTA), vol. 1 (2005), 15 pages. <http://www.ejta.org>
- [11] A. Illényi, Gy. Wersényi: Environmental Influence on the fine Structure of Dummy-head HRTFs," in Proc. 2005 Forum Acusticum, 2529-2534.
- [12] Gy. Wersényi: Evaluation of a MATLAB-Based Virtual Audio Simulator with HRTF-Synthesis and Headphone Equalization. in Proc. of 2012 ICAD12, 5 pages.
- [13] P. Minnaar, J. Plogsties, S. K. Olesen, F. Christensen, H. Møller: The Interaural Time Difference in Binaural Synthesis. 108th AES Convention Preprint 5133, Paris, 2000.
- [14] J. Nam, J. S. Abel, J. O. Smith III: A Method for Estimating Interaural Time Difference for Binaural Synthesis. 125th AES Convention Preprint 7612, San Francisco, 2008.
- [15] G. F. Kuhn: Model for the interaural time differences in the azimuthal plane. J. Acoustical Soc. Am., vol. 62(1), (1977) 157-167.
- [16] V. Larcher, J.-M. Jot: Techniques d'interpolation de filtres audio-numériques, Application à la reproduction spatiale des sons sur écouteurs. in Proc. of 1997 4th Congress of the French Soc. of Acoustics, 4 pages.
- [17] R. C. G. Smith, S. R. Price : Modelling of Human Low Frequency Sound Localization Acuity Demonstrates Dominance of Spatial Variation of Interaural Time Difference and Suggests Uniform Just-Noticeable Differences in Interaural Time Difference. 2014. PLoS ONE 9(2): e89033. doi:10.1371/journal.pone.0089033