Audio Engineering Society

Convention Paper 9766

Presented at the 142nd Convention
2017 May 20–23 Berlin, Germany

Comparison of HRTFs from a Dummy-Head Equipped with Hair, Cap and Glasses in a Virtual Audio Listening Task over Equalized Headphones

György Wersényi[1], and József Répás[1,2]

[1]*Széchenyi István University*

[2]*Óbuda University*

Correspondence should be addressed to József Répás (jozsef.repas@t-online.hu)

ABSTRACT

Head-Related Transfer Functions (HRTFs) are frequently used in virtual audio scene rendering in order to simulate sound sources at different spatial locations. The use of dummy-head HRTFs (also referred as generic sets) is often criticized because of poor localization performance, leading to e.g. lower spatial resolution, in-the-head localization, front-back reversals etc. This paper presents results of horizontal plane localization obtained by digital filter representations of dummy-head HRTFs that were recorded normally, and using additional cap, glasses and hair on the head. Results of untrained subjects over equalized reference headphones showed no significant difference among the HRTF sets despite of large magnitude differences. This method for customization of generic HRTFs fails if improvement in localization is needed.

1 Introduction

HRTFs represent the filtering effects of the outer ears, head and torso as sound waves travel from the source to the eardrums [1-5]. Through accurate recordings of the HRTFs, a limited set of filters can be designed for the left and right ear respectively. Measurements can be made on human subjects and on dummy-heads [6-8]. The latter are often referred to as non-individualized or generic sets. Usually, HRTF sets have limited spatial resolution and accuracy in time and frequency. During rendering, sound files are filtered with the HRTFs either in the time domain or in the frequency domain. According to the binaural technique, using a linearized transmission chain (including the headphone), theoretically creates a "perfect" illusion of the spatial information for the listener [9, 10]. In practice however, the limited accuracy and resolution of the HRTFs, as well as the lack of individualization and/or head-tracking, together with incomplete equalization of the headphones etc. result in decreased localization performance in contrast to free-field localization. Errors such as increased localization blur, front-back reversals, lack of externalization are well-known phenomena. One of the reasons for this is the HRTF set itself, especially in case of dummy-head recordings [11-12].

Dummy-head HRTFs are still frequently used and can be accessed in public databases [13, 14]. Some high-level programming frameworks (e.g. CSound, Pure Data) also support built-in HRTFs of dummy-heads for sound scene rendering. Head and torso simulators have the advantage of being based on long-time recordings which result in a database of HRTFs with large spatial resolution and accuracy in frequency (increased signal-to-noise ratio) [15]. On the other hand, as the data obtained reflects a single manikin, it leads to a generic set for every listener. Therefore, individually measured HRTFs are often used if high quality localization is required. Individual measurements are usually more problematic, from placing the microphones to signal presentation. Furthermore, the headphones used for playback also need to be measured and equalized individually – which is also easier using a dummy-head.

As HRTFs represent every effect of the transmission path from the source to the eardrum, even in non-reverberant free-field environments (anechoic rooms) small movements of the subject or of the microphone or other objects near the head can influence the recording. Reflections and damping effects of the sitting person's legs, shoulders, clothing and other everyday objects can also affect the HRTFs [16-22]. In our former experiments, dummy-head HRTFs were recorded with high accuracy. The effect of different haircuts, glasses, baseball caps and hats were recorded and analyzed based on spectral HRTF Differences (HRTFD) [16, 23, 24]. This analysis revealed large deviations in the magnitude of the HRTFs in given spatial directions and frequency regions. This paper presents results of localization tasks using these HRTFs over equalized headphones in search for detectable deviations also in virtual sound scene rendering as well.

2 Measurement setup

An application in MATLAB was developed to test the HRTFs. Figure 1 shows a screenshot of the GUI. Mono wave files can be loaded, plotted, processed and filtered by the HRTFs. Results can be exported to stereo wave files as single sources or dynamic movements. ITD is set by the Woodworth-formula.

Our HRTFs originate from an anechoic measurement of the Brüel&Kjaer 4128C dummy-head in 1-degree spatial resolution in the horizontal plane and 5-degree resolution in the vertical plane [15]. During measurement, HRTFs of the "naked" torso (normal set) and HRTFs with hair, glasses and baseball caps were recorded.

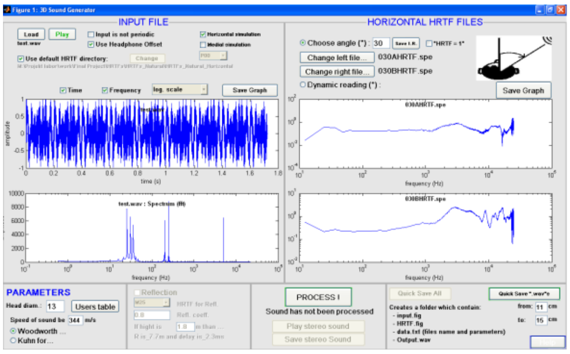

Figure 1. Screenshot of the GUI.

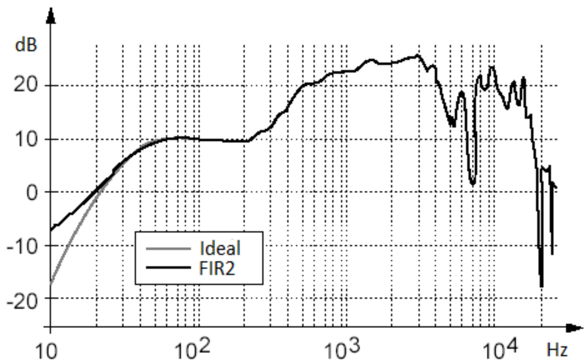

Figure 2. A 4096-tap FIR filter designed in MATLAB using the „fir2" function based on averaged measurements of the headphone's transfer function. Light grey shows the target (ideal) function, black indicates the designed FIR filter characteristics. This has to be inverted to obtain the equalizing filter.

As previously shown, repeated measurements using these objects near the head result in deviations up to 20 dB in various directions in the amplitude characteristics of the HRTFs [16, 23]. For subjective evaluation, representatives of the HRTF database were selected to emulate sound source directions in the horizontal plane only.

For the listening tasks, Sennheiser HD650 reference headphones were used. The two-channel transfer function of the headphone was measured on the same dummy-head using the B&K PULSE system. A mean transfer characteristic was calculated based on ten measurements after replacing the headphone on the head. Another MATLAB application calculated the inverse transfer function of the mean and different FIR and IIR filters were designed using MATLAB's built-in methods. For the test, a FIR-filter of 4096 taps was applied (Fig.2).

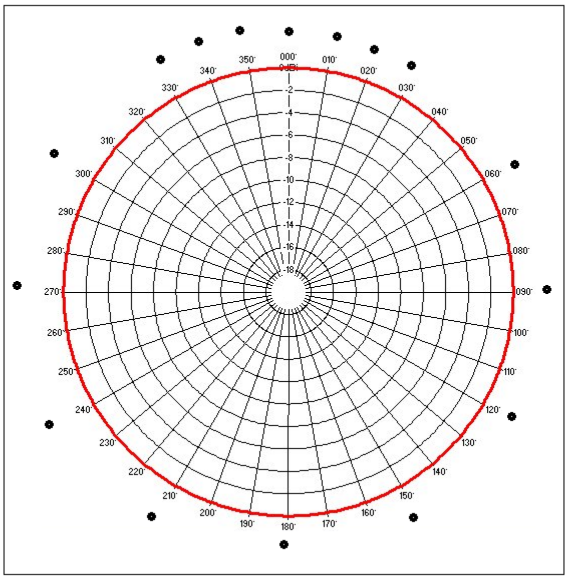

Figure 3. Simulated sound source directions in the horizontal plane (0°, 10°, 20°, 30°, 60°, 90°, 120°, 150° and 180°).

The testing application filters the input source (5 sec. white noise) with the equalization filter and the dedicated HRTF for the left and right side. The output file is exported as a stereo wave file.

30 subjects participated in the test (22 males, 8 females), most of them healthy young adults. All subjects listened to all four possible HRTF sets, where simulated directions were randomized. All sound sources were played back once, except the front and back directions (0 and 180 degrees), which were generated three and four times in the same round respectively. Thus, we used 21 test signals for each HRTF set in randomized order.

The rationale behind simulating front and back direction more frequently was to test the ratio of front-back confusions. Subjects reported their responses on a printed polar diagram by pointing with the finger (Fig.3). The possible sound source directions were not indicated on the paper.

3 Results

RESULTS OF THE LOCALIZATION TEST

Table 1 shows summarized results for all simulated source directions and HRTF sets. The left column indicates how many subjects delivered an answer at all, while the right column indicates correct hits. For example, in the case of normal HRTFs and a simulated sound source in the front (0 deg), subjects did not answer at all 33% of the time. Out of the remaining 77% of the answers, 19% was correct, 58% was incorrect.

RESULTS OF FRONT-BACK CONFUSIONS

Figure 4 shows accumulated answers in the case of front and back sources using the normal HRTF set.

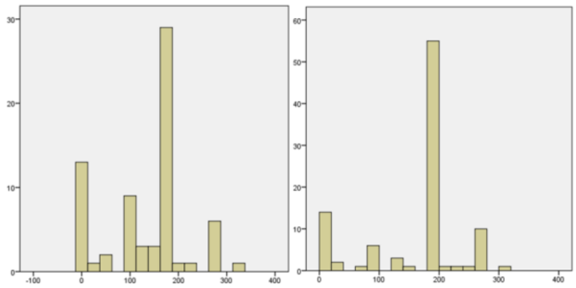

Figure 4. Results of front-back confusion using the normal HRTF set. Number of answers given by the subjects in case of a frontal source (0°, left side) and in case of a source in the back (180°, right side).

Table 2 shows all results for all HRTF sets. For example, when the sound source was rendered using normal HRTFs, the front-back error rates were 42% (front-to-back) and 15% (back-to-front). Interestingly, subjects also answered left and right (6-13%) as well as other possible directions (12-17%).

4 Discussion

Based on table 1, the localization blur is significantly worse than in free-field localization. Using the normal set, the following directions were detected with the highest accuracy: 270° (63% hits); 240° (38% hits); 210° (34 % hits); 150° (59% hits); 120° (50% hits). On the other hand, directions 20° (0%), 10° (7%), 300°(7%), 30°(4%), 350°(12%), 60°(11%) were detected with the worst accuracy. This is somewhat surprising, as these directions are in the frontal hemisphere. Due to the relatively high rate of front-back confusions (including also mirrored images to the interaural axis such as 30°-150°), sources in the back were detected more accurately. While „front" was detected with 19-31% accuracy, „back" was identified with 57-72% accuracy. The same tendency is visible e.g. for 30 and 150 degrees.

Evaluating the other sets, similar trends can be seen. Using HRTFs with hair, the most accurate directions were: 120° (67%); 240° (50); 210° (52%). The least accurate directions were: 20° (7%); 10° (0%); 30° (0%); 60° (10%); 350° (7%).

In the case of HRTFs with glasses, the most accurate directions were 150° (66%); 210° (50%); 90° (54%) and the least accurate were 20° (7%), 10° (7%), 30° (3%), 300° (7%).

In the case of HRTFs with a baseball cap, the most accurate directions were 120° (55%); 240° (55%); 90° (48%) and the least accurate were 20°(0%), 10°(8%), 60°(10%), 350°(7%).

This trend suggests high rates of front-back confusions but no significant difference among the four HRTF sets. There exist directions where one or the other HRTF set delivered better results (number of correct hits), but averaging results over all directions does not show statistically significant differences among the HRTF sets. We can conclude that despite large differences in the HRTFs introduced by these objects, dummy-head HRTFs in general cannot be personalized or adjusted effectively by applying these objects on the head. Whether this method influences individually recorded HRTFs (that is, a subject with glasses performs better with HRTFs with glasses) is an open question. One can speculate that the use of a generic set of dummy-heads is insufficient irrespective of other disturbances.

As expected, wearing a baseball cap did not introduce significant effects in the case of horizontal plane sources. The visor of the cap affects HRTFs only at higher elevations, mostly above 30 degrees. Real shadowing effects, thus, distorted high frequency components appear only around 45-60 degrees of elevation [16, 24, 25].

Evaluating front-back confusion rates with normal HRTFs shows different confusion rates in the case of front and back sources. If the source is simulated in the front, 42% of the answers indicate a backward direction (180 degrees). There is also a significant localization blur left and right from the backward direction. If the source is simulated in the back, only 15% of front answers appeared and the blur was also smaller. This is not surprising, as headphone playback systems usually suffer from increased number of detected sources in the back hemisphere, thus, back sources are easier to report correctly.

Detecting the left and right directions was not a difficult task, as 93-100% gave some answer, and correct hits ranged from 25% to 63%.

As long these objects rather influence high frequency components, and the shadowing effect of the visor is detectable above 45 degrees elevation, subjective tests in the vertical plane are put to future work.

5 Summary

A listening test was conducted with 30 untrained subjects to compare different sets of dummy-head HRTFs. HRTFs were recorded using the normal torso, and by applying different toupees, glasses and baseball caps. As long these objects influence the HRTFs significantly, psychoacoustic effects were examined in virtual horizontal plane localization tests by selecting representative HRTFs and an equalized headphone. Results revealed increased front-back reversals and no significant difference between the HRTF sets in general. Future work will include a listening test outside the horizontal plane and a second one testing other, application-dependent HRTF features (effect of reverberation, multiple sources, different headphone equalization filters etc.).

6 Acknowledgement

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 643636 "Sound of Vision".

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] Blauert: *Spatial Hearing*. The MIT Press, MA, 1983.

[2] C. I. Cheng, and G. H. Wakefield, "Introduction to Head-Related Transfer Functions (HRTFs): Representations of HRTFs in Time, Frequency, and Space," *J. Audio Eng. Soc.,* vol. 49, no. 4, pp. 231-249, 2001.

[3] H. Møller, M. F. Sorensen, D. Hammershøi, and C. B. Jensen, "Head-Related Transfer Functions of human subjects," *J. Audio Eng. Soc.,* vol. 43, no. 5, pp. 300-321, 1995.

[4] F. Wightman, D. Kistler, "Measurement and validation of human HRTFs for use in hearing research" *Acta acustica united with Acustica,* vol. 91, pp. 429-439, 2005.

[5] W. M. Hartmann, "How we localize sound," *Physics Today*, vol. 52, no. 11, pp. 24-29, November 1999.

[6] H. Møller, D. Hammershøi, C. B. Jensen, and M. F. Sorensen, "Evaluation of artificial heads in listening tests," *J. Audio Eng. Soc.,* vol. 47, no. 3, pp. 83-100, 1999.

[7] H. Møller, M. F. Sorensen, C. B. Jensen, and D. Hammershøi, "Binaural Technique: Do We Need Individual Recordings?," *J. Audio Eng. Soc.,* vol. 44, no. 6, pp. 451-469, 1996.

[8] P. Minnaar, J. Plogsties, F. Christensen, "Directional resolution of Head-related transfer functions required in binaural synthesis" *J. Audio Eng. Soc,* vol. 53, no. 10, pp. 919-929, 2005

[9] H. Møller, "Fundamentals of binaural technology," *Applied Acoustics,* vol. 36, pp. 171-218, 1992.

[10] P. Minnaar, S. K. Olesen, F. Christensen, and H. Møller, "Localization with Binaural Recordings from Artificial and Human Heads," *J. Audio Eng. Soc.,* vol. 49, no. 5, pp. 323-336, 2001.

[11] P. F. Hoffmann, and H. Møller, "Audibility of Direct Switching Between Head-Related Transfer Functions," *Acta Acustica united with Acustica*, vol. 94, no. 6, pp. 955-964, 2008.

[12] 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, no. 10, pp. 904-917, 2001.

[13] http://interface.cipic.ucdavis.edu/CIL_html/CIL_HRTF_database.htm

[14] IRCAM, LISTEN HRTF database: http://recherche.ircam.fr/equipes/salles/listen/

[15] Gy. Wersényi, "Measurement system upgrading for more precise measuring of the Head-Related Transfer Functions," in *Proc. of InterNoise2000*, Nice, France, 2000, pp. 1173-1176.

[16] Gy. Wersényi, and A. Illényi, "Differences in Dummy-Head HRTFs Caused by the Acoustical Environment Near the Head," *Electronic Journal of "Technical Acoustics" (EJTA)*, vol. 1, 15 pages, 2005. http://www.ejta.org

[17] P. F. Hoffmann, and H. Møller, "Some observations on sensitivity to HRTF magnitude," *J. Audio Eng. Soc.*, vol. 56, no. 11, pp. 972-982, 2008.

[18] K. A. J. Riederer, *Head-related transfer function measuremnts,* Master Thesis, Helsinki University of Technology, 1998.

[19] K. A. J. Riederer, HRTF analysis: Objective and subjective evaluation of measured head-related transfer functions, Dissertation, Helsinki University of Technology, Espoo, 2005.
[20] B. E. Treeby, J. Pan, and R. M. Paurobally, ”The effect of hair on auditory localization cues,” *J. Acoustical Soc. Am.*, vol. 122, no. 6, pp. 3586-3597, December 2007.
[21] B. E. Treeby, J. Pan, and R. M. Paurobally, ”An experimental study of the acoustic impedance characteristics of human hair,” *J. Acoustical Soc. Am.*, vol. 122, no. 4, pp. 2107-2117, October 2007.
[22] P. F. Hoffmann, and H. Møller, ”Audibility of Differences in Adjacent Head-Related Transfer Functions,” *Acta Acustica united with Acustica*, vol. 94, no. 6, pp. 945-954, 2008.
[23] A. Illényi, and Gy. Wersényi, ”Environmental Influence on the fine Structure of Dummy-head HRTFs,” in *Proc. of the Forum Acusticum 2005*, Budapest, 2005, pp. 2529-2534.
[24] A. Illényi, and Gy. Wersényi, ”Evaluation of HRTF data using the Head-Related Transfer Function Differences,” in *Proc. of the Forum Acusticum 2005*, Budapest, 2005, pp. 2475-2479.
[25] B. F. Katz, “Acoustic absorption measurement of human hair and skin within the audible frequency range,” *J. Acoustical Soc. Am.*, vol. 108, no. 5, pp. 2238-2242, 2000.

	Normal		Hair		Glasses		Baseball cap	
Deg	**answer (%)**	**hit (%)**	**answer (%)**	**hit (%)**	**answer (%)**	**hit (%)**	**answer (%)**	**hit (%)**
0	77%	19%	78%	31%	77%	26%	84%	24%
10	93%	7%	90%	0%	93%	7%	87%	8%
20	100%	0%	97%	7%	97%	7%	90%	0%
30	93%	4%	100%	0%	97%	3%	97%	17%
60	93%	11%	100%	10%	97%	17%	97%	10%
90	93%	25%	97%	48%	93%	54%	97%	48%
120	100%	50%	100%	67%	97%	41%	97%	55%
150	97%	59%	100%	50%	97%	66%	97%	38%
180	80%	57%	79%	72%	78%	59%	78%	60%
210	97%	34%	97%	52%	93%	50%	90%	48%
240	97%	38%	100%	50%	97%	41%	97%	55%
270	100%	63%	100%	40%	97%	34%	97%	28%
300	97%	7%	100%	13%	97%	7%	97%	17%
330	93%	18%	93%	11%	97%	24%	97%	17%
340	97%	21%	97%	24%	97%	24%	93%	18%
350	87%	12%	90%	7%	97%	14%	97%	7%

Table 1. Results for each simulated direction using normal HRTFs and with hair, glasses and baseball cap.

	Hits	***front-back***	***back-front***	***270***	***90***	***other***
Normal						
180	***57%***	***-***	***15%***	***10%***	***6%***	***12%***
0	***19%***	***42%***	***-***	***9%***	***13%***	***17%***
Hair						
180	***72%***	***-***	***2%***	***1%***	***12%***	***14%***
0	***31%***	***41%***	***-***	***11%***	***14%***	***3%***
Glasses						
180	***59%***	***-***	***13%***	***3%***	***13%***	***12%***
0	***26%***	***49%***	***-***	***1%***	***22%***	***1%***
Baseball cap						
180	***60%***	***-***	***24%***	***1%***	***6%***	***9%***
0	***24%***	***59%***	***-***	***5%***	***3%***	***9%***

Table 2. Summarized results in case of front and back sources.