COMPUTATIONAL QUALITY ASSESSMENT OF HRTFS

Klaus A J Riederer
Laboratory of Computational Engineering
Laboratory of Acoustics and Audio Signal Processing
Helsinki University of Technology (HUT)
P.O. Box 9400, 02015 HUT, Finland
Tel: +358 9 451 5328; fax +358 9 451 4830
e-mail: Klaus.Riederer@hut.fi, URL: www.lce.hut.fi/~kar

ABSTRACT

The verification of quality in head-related transfer function (HRTF) measurement data holds a complex problem field, involving acoustic (background noise), physical (head movement, different head shape) and electrical (DC-offset) deviations. The disrupted measurements may be difficult to separate from the correct ones, because HRTFs are inherently highly dependent on the person, direction, and frequency. Since there exists no “absolute truth” what a HRTF should be like, deviated measurements can be assessed from the typical HRTF shape, i.e., the common trend of the measured HRTFs. This paper presents an approach to calculate an objective deviant measure of HRTFs by DSP means. The perceptually justified deviance index is applied for the evaluation of a HRTF measurement system and a HRTF database.

1 MODELING OF HRTFS

The human auditory system perceives the spatial information of a sound source by detecting interaural time and level differences (ITDs and ILDs), as well as by the spectral cues caused by the human body (Blauert 1997). The interaction of these matters is denoted as head-related transfer functions (HRTFs) that involve measured (or modeled) responses of a sound source in a free-field space to a point in the ear canal (Møller 1992). The HRTFs are measured as impulse responses (usually ca. 100-200 ms long) and then synthesized by digital filters in order to produce realistic 3-D sound over headphones or loudspeakers.

Usually, in practice mathematical modeling is used to simplify massive HRTF data storage and its control. Moreover, the data reduction may also be justified based on the human perceptual accuracy. The peripheral resolution of a human is poorer at high frequencies because of the wider critical bandwidth, or auditory filter width. Also various psychoacoustic experiments show that the localization capabilities depend on the sound direction (see, e.g., Blauert 1997).

Different methods of HRTF modeling have been presented in the literature. For example, spatial feature extraction (Chen *et al.* 1993) and spherical harmonic analysis (Evans *et al.* 1997) include both magnitude and phase information, and their continuous representations allow interpolation of unmeasured directions. Kistler and Wightman (1992) modeled the HRTF magnitude by principal component analysis (PCA) of measured HRTFs. Their first five basis vectors (principal components) explained 90% of total variance of the original data. The perceptual tests suggested that PCA captured nearly all the relevant details in the HRTFs. Motivated by this study, the author also used PCA as a starting point.

2 PRINCIPAL COMPONENT ANALYSIS OF HRTFS

In general, the magnitude responses of HRTFs are fairly flat up to 1-2 kHz, but after that the magnitudes vary vigorously, with a peak-to-peak value greater than 40 dB and very different person-dependent fine structure (Riederer 1998ab). This suggests that pre-processing, e.g., subtracting subject-dependent and direction-independent information from the data is relevant, before a common trend can be estimated.

The database for the PCA (Riederer 1998ab) consists of both ear HRTF magnitude responses from 55 subjects. First auditory-based frequency warping, i.e., optimal BARK warping is used (Smith and Abel 1995). The amount of input data points is considered, so that only notable/remarkable deviations are noted — an amount of 128 points was found to be the best suitable. The responses from each direction and for each ear are investigated at a time; hence the input data matrix has 128 rows (p) and 55 (n) columns.

The warped frequency responses up to the Nyquist frequency (0 - 24000 Hz) are centered by a subject-mean, i.e., by subtracting the mean value of a subject (per ear and direction). The values are then scaled, dividing them by a factor of $\sqrt{p-1}$. The removal of the mean value diminishes the unwanted subject-dependent and direction-independent effects on the input data. The first 2-7 principal components (PCs) explain 95% of the total variance, depending slightly on the direction. As anticipated, the first PC explains clearly the most of the original data, 80-90%, of the total variance.

3 DEVIANCE INDEX

The principal component reconstruction equals to a weighted summation over the input data, similar as in Fourier synthesis. If the HRTF reconstruction is done by applying the first principal component (the first basis vector) and the corresponding (individual) weights, only a minimum amount of individual features is introduced in the obtained HRTF model (PCA_1). Therefore, PCA_1 represents a "reference" HRTF, i.e., the common trend of all the measured HRTFs (per ear and direction), see Figs 1-3. (In this paper only right ear responses are shown due to length limitations.)

An *error-count* (*ERR*) is obtained as a square error over all the frequency points between the PCA_1 and the original measured HRTF, divided (normalized) by the total energy of the PCA_1 :

$$ERR_{person,angle,ear} = \frac{\sum_{n} (PCA_{1,person,ear,n} - HRTF_{person,ear,n})^2}{\sum_{n} PCA_{1,person,ear,n}^{\;2}}$$

Then, *ERR* is scaled so that a suitable range is obtained for the *deviance* (*DEV*):

$$DEV_{person,angle,ear} = 1 - \tanh\left(\alpha_{angle,ear} * ERR_{person,angle,ear}\right)$$

where the scaling factor has been chosen experimentally as

$$\alpha_{angle,ear} = \frac{\text{arctanh}(0.5)}{mean\left(ERR_{person,angle,ear}\right)}.$$

Hence, the value of the deviance index is [0,1] limited. A measured HRTF that differs strongly from the common trend of all the HRTFs has a low DEV, close to 0. It should be understood that the deviance index is an implicit measure, showing relative behavior of HRTFs — it does not show the explicit "goodness" of HRTFs.

4 STATISTICAL ANALYSIS OF HRTF DATABASE

The calculated deviance indexes can be illustrated by a number of means showing different aspects. The most illustrative way to study the deviant measurements is to assign a certain threshold, DEV_{th} , and indicate the DEVs below it. Fig. 4 indicates with a cross sign (+) all right ear measurements, for which $DEV_{th} < 0.2$. The azimuth angles are plotted on the vertical axis and the number of the test person on the horizontal axis. Crosses that form vertical lines reveal the most deviant test subjects, i.e., here the subjects numbered 2 and 54. This result is anticipated, since these subjects were dummy heads, whose ear canal's are wider and shorter and more cylinder-shaped than the human corresponding. Therefore, the dummy heads constitute a different HRTF structure than human heads (see Figs 1 and 3).

The distribution of the deviance indexes reveals the overall quality of the HRTF measurement system. The quality of the system would be questionable, if there would be certain a) measurement angle(s), b) elevation(s) or c) azimuth(s) giving much lower DEVs than in general. In Fig. 4, "faulty" azimuths would be clearly indicated by crosses forming horizontal lines. In order to examine the DEV distribution, median plots with upper and lower quartiles are presented in Fig. 5.

Fig. 5a shows the median values (0.55..0.8) per elevation (37 azimuths averaged per person). The variation between the subjects is minor (< 0.2) and somewhat more for the 90° elevation (~ 0.3; slightly less for the left ear responses). Most likely this measurement direction is more susceptible to asymmetrical measurement positions in the median plane, i.e., pivoting (sideward tilting) of the head, than the other directions. An ITD analysis shows that this pivoting is minor in total (Riederer 1998a).

Fig. 5b shows the median values per azimuth angle of all subjects and their averages over all elevation angles. Now, the median values vary much less (0.7..0.8) than in the prior elevation case, which is also visible in Fig. 4. The result is evident, since there are a lower amount of measurements analyzed in this case (seven elevations averaged per person). There is a minor variation between different azimuth angles, slightly higher DEVs are obtained at ipsilateral directions (the same is observed for the left ear data). The reason for this is that the HRTF shape is smoother in the head shadow directions, which causes slightly smaller ERR values.

Figs 4, 5a and 5b indicate that there does not exist any notable systematic mechanical/acoustical faults in the measurement system, i.e., the high quality and stability of the system; yielding the same result as in (Riederer 1998ab).

Fig. 5c reveals the large variation between the test subjects; the maximum deviance is noted with subject numbers 2 and 54 and minimum with subjects 10 and 39. These inter-individual results give an indication for a clustering analysis of the HRTF data.

5 DISCUSSION

A deviation from the common trend may be caused by several (possibly coincidental) reasons: acoustic, electrical or electrical faults (Riederer and Karjalainen 1998). Fortunately, they have different effect to the frequency responses (HRTFs), e.g., reflections are time-delayed unwanted signals, DC-offsets show deviations at low frequencies. However, the physical faults are more problematic: vivid head movement cause strong deviations from the common HRTF trend above 1..2 kHz, but also a very atypical head shape of a test subject would lead to the same situation — without being any measurement fault (!). Furthermore, HRTFs contain person-dependent resonance structures at high frequencies (above ca. 7 kHz), that might have somewhat generic shape but are usually shifted in frequency. This makes it very difficult to both find a common HRTF structure or to observe imperfections in the given HRTF.

The method presented in this paper proved the high quality of the HRTF measurement system but does not show explicitly degraded HRTF data per person. For the latter purpose, more sophisticated method analyzing also the band-limited frequency response and the time-domain response is necessary. However, the most important is the perceptual effect: HRTFs are mostly applied for 3-D audio, hence listening tests are unavoidable when assessing the quality of HRTF data.

ACKNOWLEDGEMENTS

The author wishes to thank Dr.Tech. Jyri Huopaniemi (Nokia Research Center), M.Sc. Matti Hämäläinen (Nokia Research Center), Dr.Tech. Vesa Välimäki (HUT), Prof. Matti Karjalainen (HUT), M.Sc. Timo Kostiainen (HUT) and M.Sc. Harri Lappalainen (HUT) for stimulating talks and giving good ideas for the research.

REFERENCES

Blauert J., 1997. *Spatial Hearing. The Psychophysics of Human Sound Localization.* Revised Edition, MIT Press, Cambridge, Massachusetts, USA, 494 p.

Møller, H. 1992. Fundamentals of binaural technology. *Applied Acoustics*, vol. 36, pp. 171–218.

Riederer K. A. J, 1998a. *Head-related transfer function measurements*. Master's Thesis. Helsinki University of Technology, Laboratory of Acoustics and Audio Signal Processing, Espoo, Finland, 134 p.

Riederer K. A. J. and Karjalainen M., 1998. DSP aspects of HRTF measurements. In Proceedings of *IEEE Signal Processing Symposium Norsig 98*, Vigsø, Denmark, June 8-11, 1998.

Riederer K. A. J, 1998b. Repeatability analysis of head-related transfer function measurements, *105th Audio Engineering Society Convention*, San Francisco, Sept. 26-29. Preprint no. 4846.

Smith, J. O. S. III and Jonathan S. A., 1995. The bark bilinear transform. Proc. *IEEE ASSP Workshop on Applications of Signal Processing to Audio and Acoustics*, Mohonk Mountain House, New Paltz, New York, October 15-18, 1995. Session 8.6.

Chen J., Veen Van B. D. and Hecox K. E., 1993. A spatial feature extraction and regularization model for virtual auditory display. In *Proc. 1993 IEEE Int. Conf. Acoust., Speech, and Signal Processing (ICASSP'93)*, Minneapolis, Minnesota, USA, vol. I, pp. 129–132.

Kistler D. J. and Wightman W. L., 1992. A model of head-related transfer functions based on principal components analysis and minimum-phase reconstruction. *J. Acoust. Soc. Am.*, **91**, 3, pp. 1637–1647.

Evans M.J., Angus J.A.S., Tew A.I. (1997). Spherical harmonic spectra of head-related transfer functions, *103th Audio Engineering Society Convention*, New York. Preprint no. 4571.

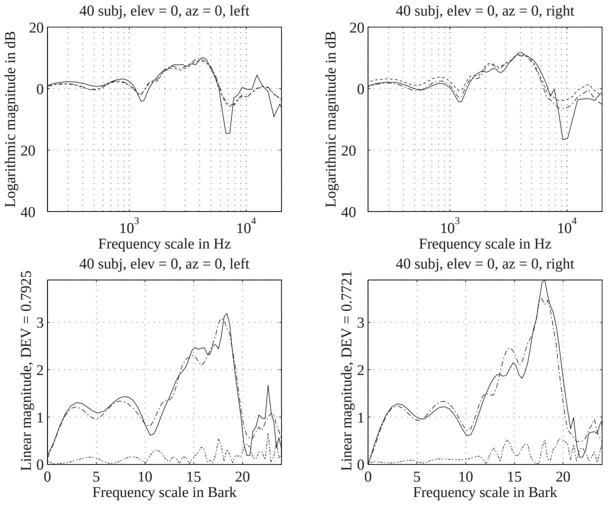


Figure 1. PC-1 reconstructed HRTFs (dash dot); mean of all the measured HRTFs (dash) is power-fitted to the measured HRTF (solid), the dotted curve in the lower figures illustrates the point-to-point difference of the measured and reconstructed HRTFs. Deviance (DEV) relates to the surface area that the blue curve makes with the horizontal axis. In this figure only a slight deviation from the common trend is noted, except at the antiresonances at high frequencies.

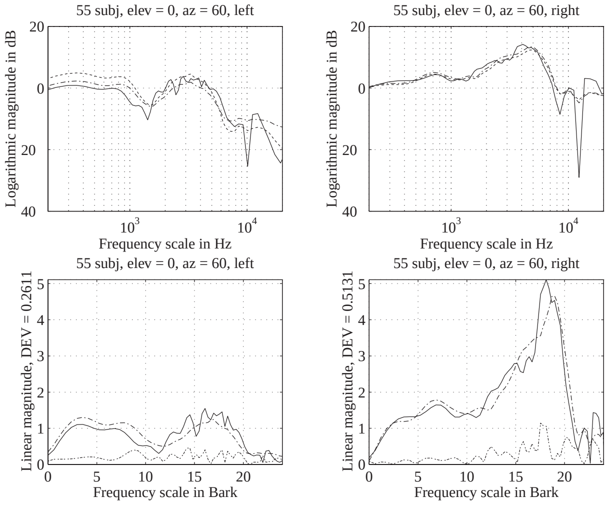
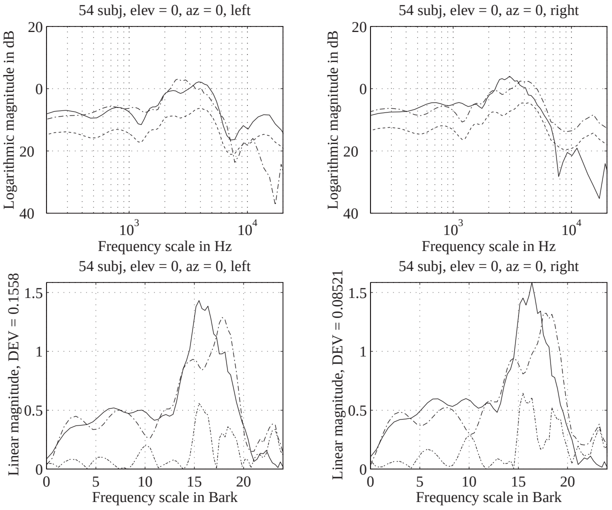


Figure 2. In the left ear rather high deviance (also at low frequencies), for the right ear only at the antiresonancies at high frequencies.


54 subj, elev = 0, az = 0, left

54 subj, elev = 0, az = 0, right

54 subj, elev = 0, az = 0, left

54 subj, elev = 0, az = 0, right

Figure 3. HRTF and DEV plot for the test subject number 2 (Cortex MKII), same measurement angle as in Fig. 1.

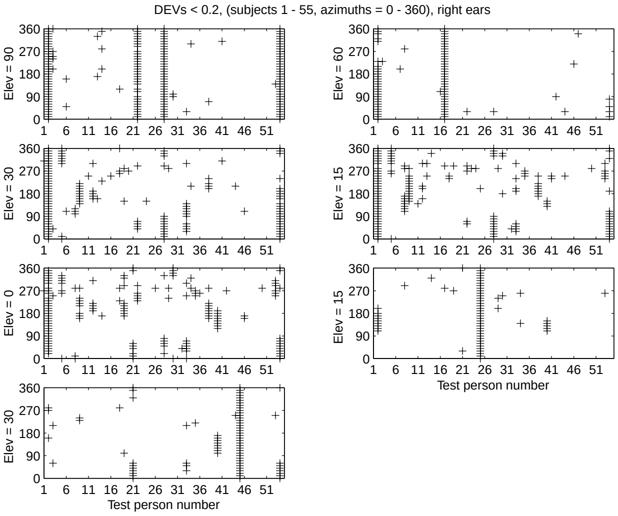
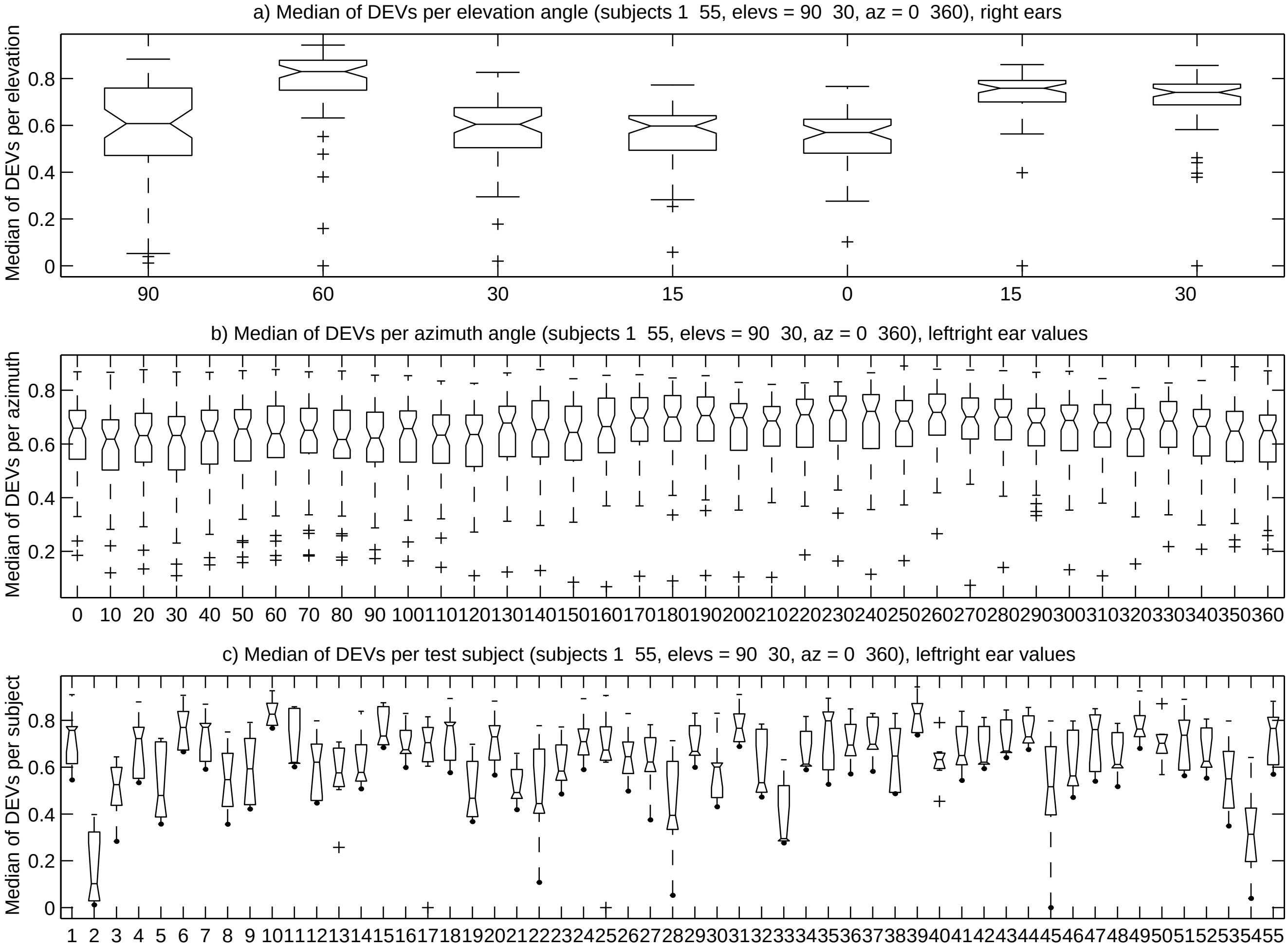

DEVs < 0.2, (subjects 1 - 55, azimuths = 0 - 360), right ears

Figure 4. Right ear HRTFs, whose DEVs are below 0.2.


a) Median of DEVs per elevation angle (subjects 1 55, elevs = 90 30, az = 0 360), right ears

b) Median of DEVs per azimuth angle (subjects 1 55, elevs = 90 30, az = 0 360), leftright ear values

c) Median of DEVs per test subject (subjects 1 55, elevs = 90 30, az = 0 360), leftright ear values

Figure 5. Median values for right ear HRTF measurements; a) per elevation angle b) per azimuth angle, c) per test subject. The lower and upper lines of the "box" indicate the 25% and 75% percentiles of the sample.