

Individualized binaural reproduction using a virtual artificial head

Eugen Rasumow¹, Matthias Blau¹, Simon Doclo², Martin Hansen¹, Steven van de Par²,
Dirk Püschel³, Volker Mellert²

¹ *Institut für Hörtechnik und Audiologie, Jade Hochschule, D-26121 Oldenburg*

² *Carl von Ossietzky Universität Oldenburg, D-26129 Oldenburg*

³ *Akustik Technologie Göttingen, D-37073 Göttingen*

Motivation

In binaural recordings, spatial information is usually captured using so-called artificial heads, which are replica of real human heads with ear microphones and average anthropometric geometries. Because of their non-individual character, such recordings often entail perceptual deficiencies (front-back confusion, internalization etc.). Alternatively, individual head-related transfer functions (HRTFs) can be re-synthesized using a microphone array in conjunction with the *filter and sum* method and appropriate digital filters (referred to as virtual artificial head, VAH). Its main advantage over traditional artificial heads is the possibility to adapt one recording to many individual HRTFs in the binaural reproduction by appropriate modification of the array's directional pattern. In this contribution, the realization of a VAH as a planar microphone array with 24 microphones is presented. The VAH is perceptually evaluated against traditional artificial heads and original HRTFs for different directions of incidence in the horizontal plane.

HRTFs

Prior to the re-synthesis of individual ear signals, individual HRTFs were measured in the horizontal plane for six participating subjects who had extensive experience with this kind of psychoacoustical experiments. HRTFs were measured using the blocked ear method with the microphones (Knowles FG-23329 miniature electret microphones) embedded in foam earplugs (cf. [2]). All HRTFs were truncated in the time domain to 256 samples (≈ 6 ms at a sampling frequency of $f_s = 44100$ Hz) with a tapered Hann-window with a descending flank of 50 samples. Based on the findings from [5], the individual HRTFs were smoothed in the frequency- and spatial domains: The phase responses of individual HRTFs were substituted by linear phases for frequencies $f > 1$ kHz, which enabled complex smoothing of the HRTFs into constant relative bandwidths of $B_W = \frac{1}{5}$ -octaves in the frequency domain. Furthermore, spatial notches of individual directivity patterns (HRTFs for a fixed frequency as a function of direction) were levelled out such that the dynamic range of the directivity patterns never exceeded 29 dB at any frequency.

Headphone transfer functions (HPTFs)

The individual HPTFs were measured right after measuring the HRTFs with the microphones still left in place. When measuring the HPTFs, subjects were instructed to reposition the headphone ten times to various realistic positions which successively yielded ten different individ-

ual HPTFs. The individual HPTF resulting in the smallest dynamic range for frequencies $300 \text{ Hz} \leq f \leq 16000 \text{ Hz}$ was inverted according to the method given in [1] with $\beta = 0.01$ after adjusting the rms-level of the individual HPTFs to -30 dB FS. This procedure was applied to ensure that the quality of the various equalizations is as constant as possible across all different devices/conditions.

Virtual artificial head (VAH)

Using the *filter and sum* method in conjunction with an appropriate microphone array and associated filters, it is possible to create certain desired frequency dependent directivity patterns. Given the motivation to approximate the individual HRTFs, individual filter coefficients for the VAH may be optimized by minimizing a chosen cost function between the desired HRTF and the re-synthesized directivity pattern. In this study, a least squares cost function in conjunction with a constraint on the mean white noise gain (WNG_m) according to the approach in [4] was used. It is worth noting here, that the traditional definition of the white noise gain (WNG) needs to be adapted in order to consider all desired directions Θ instead of a single steering direction Θ_0 . Furthermore,

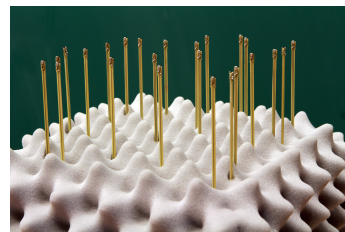


Figure 1: Used microphone array with 24 sensors with the two dimensional topology according to [3]

based on the findings from [4], a planar microphone array with 24 sensors (cf. Figure 1) composed of MEMS microphones (Analog Devices ADMP 504 Ultralow Noise Microphone) was used for re-synthesizing the individual ear signals for discrete directions in the horizontal plane with an angular resolution of $\Delta\Theta = 15^\circ$. The filter coefficients were calculated according to the procedure in [4] with a desired $\text{WNG}_m = 0$ dB.

Methods

The aim of this study is to compare the appraisals of individualized re-syntheses using the VAH with the appraisals when using traditional artificial heads. To this end, bursts of pink noise ($200 \text{ Hz} \leq f \leq 16000 \text{ Hz}$) arriving from three directions in the horizontal plane ($\Theta = 90^\circ$ left, $\Theta = 0^\circ$ frontal and $\Theta = 225^\circ$ back-right) were

recorded with the VAH and with three traditional artificial heads (*G.R.A.S. KEMAR* (Kemar), *Head Acoustics HMSII.2* (Head) and a custom-made artificial head (SP)) in an anechoic chamber. The headphone signal was calibrated individually to have 78 dB SPL (*G.R.A.S.* type 43AA artificial ear) for the ipsilateral side and $\Theta = 90^\circ$. Additionally, a continuous white noise signal of 40 dB SPL was added to the noise bursts in order to mask the individually-colored sensor noise of the various devices. In a further condition (HRTF), the test signals were filtered with individual HRTFs. All test signals were equalized with the HPTFs measured with the associated device (Kemar, Head or SP) or the individually measured HPTFs (HRTF and VAH). These five test signals were evaluated by the subjects with reference to a presentation of the test signal presented via loudspeakers (reference condition), where subjects could switch between the test condition (presented in a hidden, randomized order) and the reference condition by pushing a toggle-button attached to the headphones. Subjects were instructed to evaluate the perceptual qualities of the test conditions regarding *localization*, *spectral coloration* and *overall performance* in three subsequent experiments with respect to the reference condition on an English category scale ranging between *bad*, *poor*, *fair*, *good* and *excellent*. Each condition was evaluated three times in a randomized order.

Results

The aggregated results (over all subjects and retests) with regard to the perceived *localization* are illustrated in Figure 2 on the y-axis as boxplots and the tested conditions on the x-axis with the tested directions separated in three boxes. The best evaluations result for the HRTF-condition followed by the individualized re-syntheses using the VAH. The evaluations for the traditional artificial heads (Kemar and Head) vary considerably over subjects, on average being below the HRTF- and VAH-evaluations, with the worst evaluations for SP.

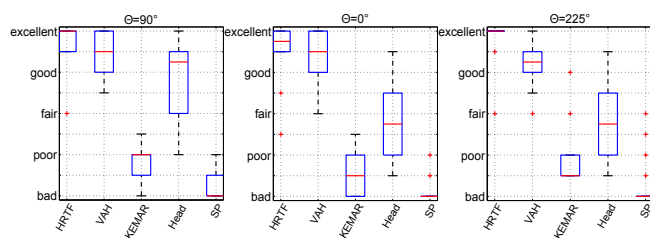


Figure 2: Aggregated evaluations (y-axis) for all subjects regarding the perception of *localization* for the five tested methods on the x-axis

The aggregated evaluations regarding *spectral coloration* are illustrated in Figure 3 analogously to Figure 2. Again, the HRTF- and the VAH-condition are evaluated best. Primarily for the frontal direction, the evaluations of the VAH re-syntheses are evaluated as *excellent*, equivalently to the HRTF condition. The evaluations of the traditional artificial heads again vary clearly over subjects, with the averaged evaluations being at or below *fair* and with the worst evaluations for SP.

The aggregated evaluations with regard to the *overall*

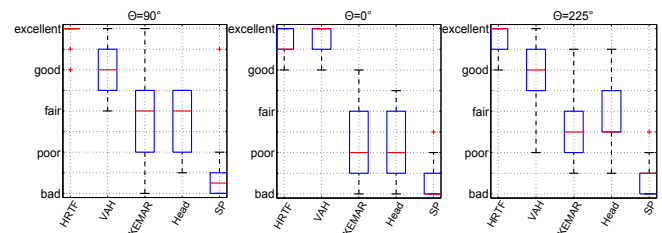


Figure 3: Aggregated results for all subjects regarding the perception of *spectral coloration* for the five tested methods

performance are illustrated in Figure 4. Consistently, the HRTF- and VAH-condition were evaluated best, especially for the frontal direction. For the other directions the mean evaluations of the VAH conditions are at or better than *good*, whereas the mean evaluations of the traditional artificial heads are at or below *fair*.

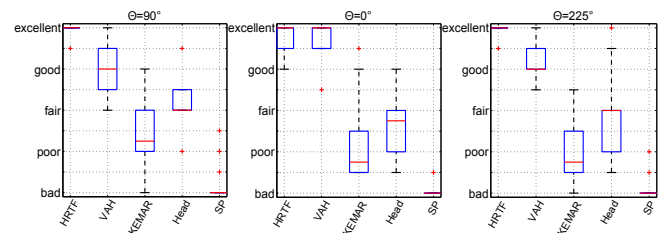


Figure 4: Aggregated results for all subjects regarding the *overall performance* for the five tested methods

Conclusion

In sum, the gathered evaluations confirm the validity of the concept of re-synthesizing HRTFs using a VAH. The results emphasize the importance of individualization for binaural re-syntheses. Further evaluations of non-optimized directions (i.e. direction in between the optimized directions) need to be investigated in order to guarantee an appropriate re-synthesis of all directions in the horizontal plane.

References

- [1] O. Kirkeby and Philip A. Nelson, "Digital filter design for inversion problems in sound reproduction," *J. Audio Eng. Soc.*, vol. 47, no. 7/8, pp. 583-595, 1999
- [2] S. Raufer, E. Rasumow, and M. Blau. HRTF- measurements with earmolds and conventional ear plugs - a comparison. In *Proceedings of AIA-DAGA 2013, Merano (Italy)*, pages 1320-1321, 2013
- [3] E. Rasumow, M. Blau, M. Hansen, S. Doclo, S. van de Par, V. Mellert, and D. Püschel, "Robustness of virtual artificial head topologies with respect to microphone positioning errors," in *Proc. Forum Acusticum, Aalborg, Aalborg*, 2011, pp. 2251-2256
- [4] E. Rasumow, M. Blau, M. Hansen, S. Doclo, S. van de Par, V. Mellert, and D. Püschel. "The impact of the white noise gain (WNG) of a virtual artificial head on the appraisal of binaural sound reproduction". In *Proc. of the EAA Joint Symposium on Auralization and Ambisonics, Berlin, Germany, 3-5 April 2014*, 2014
- [5] E. Rasumow, M. Blau, M. Hansen, S. van de Par, S. Doclo, V. Mellert, and D. Püschel, "Smoothing individual head-related transfer functions in the frequency and spatial domains", *Journal of the Acoustical Society of America*, 135(4), pp. 2012-2025, April 2014.