

Proceedings of Meetings on Acoustics

Volume 19, 2013

<http://acousticalsociety.org/>

ICA 2013 Montreal
Montreal, Canada
2 - 7 June 2013

Signal Processing in Acoustics

Session 4aSP: Sensor Array Beamforming and Its Applications

4aSP6. Least squares versus non-linear cost functions for a virtual artificial head

Eugen Rasumow*, Matthias Blau, Simon Doclo, Martin Hansen, Steven Van de Par, Dirk Püschel and Volker Mellert

*Corresponding author's address: Institut für Hörtechnik und Audiologie, Jade Hochschule Oldenburg, Oldenburg, 26121, Niedersachsen, Germany, eugen.rasumow@jade-hs.de

In order to take into account spatial information into binaural recordings it is common practice to use so-called artificial heads. Disadvantageously artificial heads are inherently non-individual and bulky devices. Alternatively, the individual frequency-dependent directivity pattern of human head related transfer functions (HRTFs) can also be approximated by a microphone array with appropriate filters (Rasumow et al., 2011). Such a setup may be referred to as a virtual artificial head (VAH). The filters for the application of the VAH can be derived by minimizing a narrowband cost function including regularization constraints. As a first approach, it is appropriate to apply a least-squares cost function. The major advantage is its closed form solution (cf. Rasumow et al., 2011), whereas from a psychoacoustical point of view, it seems more reasonable to minimize the dB-error instead. The latter cost function must, however, be minimized iteratively. We propose a minimization procedure for and present first results regarding the subjective appraisal of binaural filters derived using both cost functions. Future work includes the extension of this work to binaural cost functions.

Published by the Acoustical Society of America through the American Institute of Physics

1. INTRODUCTION

The use of so-called artificial heads, which are a replica of real human heads, has become common practice for binaural recordings today. In this way the signals at the ears receive characteristic spatial information, which encompasses interaural time and level difference cues, but also spectral cues due to the shape of the pinna, for instance. Yet, artificial heads are bound to non-individual (average) anthropometric geometries and are most often implemented as bulky devices.

Alternatively, the desired spatial filtering associated with individual head-related transfer functions (HRTFs) can be approximately re-synthesized by a microphone array with appropriate digital filtering (cf. Tohtuyeva and Mellert (1999), Atkins (2011), Kahana *et al.* (1999) and Rasumow *et al.* (2011)). Such a setup may be referred to as a virtual artificial head (VAH, cf. Fig. 5). Virtual artificial heads are much more flexible than real artificial heads, since, for instance, the filters can be adjusted to match an individual set of HRTFs after a recording has been made.

The re-synthesized directivity pattern of the VAH, *inter alia*, depends on the number and the topology of the microphones in the array but also on the design procedure for the filter coefficients. Consequently, the used cost function for optimizing the filter coefficient may have a major influence on the re-synthesized directivity pattern and, eventually, on the subjective impression. In this article, the re-synthesis using a least-squares and a non-linear cost function are compared and evaluated in a subjective A/B-comparison. Moreover, the advantages and disadvantages of the corresponding cost functions are discussed.

2. CALCULATION OF FILTER COEFFICIENTS

The VAH consists of N spatially distributed microphones, with its frequency- and direction-dependent output $H(\omega, \Theta)$ being the sum of the N microphone signals filtered with the filter coefficients $\mathbf{w}(\omega)$. The filter coefficients $\mathbf{w}(\omega)$ are computed by minimizing a cost function J that is a measure of the difference between the desired directivity pattern $D(\omega, \Theta)$ and the re-synthesized directivity pattern $H(\omega, \Theta)$ (cf. Fig 1). In general, the cost function can either be minimized for all frequencies simultaneously (broadband optimization) or separately for each frequency bin (narrowband optimization). In the following computations only narrowband optimization will be considered because filters can be optimized independently for each frequency.

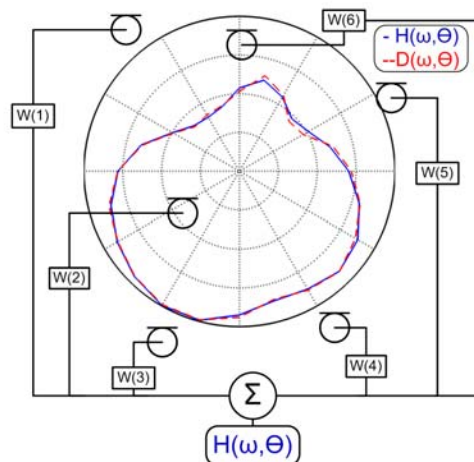


FIGURE 1: The sum of the filtered microphone signals $H(\omega, \Theta)$, which is the output of the VAH, is adjusted to approximately re-synthesize the desired directivity pattern $D(\omega, \Theta)$ by optimizing the filter coefficients $\mathbf{w}(\omega)$.

The chosen cost function not only determines the error to be minimized but also affects the optimization method. A least-squares cost function, for instance, has the advantage of a closed form solution. Yet, a regularization constraint is generally needed to ensure robustness against gain, phase and position errors of the microphones within the VAH (cf. Rasumow *et al.* (2011)). The calculation of regularized filter coefficients using a least-squares cost function is described in section 3. Despite the advantages of a least-squares cost function, a psychoacoustically motivated approach would be to minimize a logarithmic error instead. Consequently, a regularized non-linear cost function is presented in section 4. In contrast to the least-squares cost function, this non-linear cost function needs to be minimized using an iterative gradient descent method.

In order to evaluate the subjective advantages and disadvantages of both cost functions, a third cost function is introduced, which merges both cost functions with a variable blending factor. In this way it becomes possible to evaluate the subjective differences of the applied cost functions by varying the blending factor.

3. LEAST-SQUARES COST FUNCTION

Consider the desired directivity pattern $D(\omega, \Theta)$ depending on the frequency ω and the azimuthal angle Θ . The re-synthesized directivity pattern of the VAH, $H(\omega, \Theta)$ is the sum of the filtered microphone signals, which can be written as

$$H(\omega, \Theta) = \mathbf{w}^H(\omega) \mathbf{d}(\omega, \Theta) . \quad (1)$$

The N -dimensional vector $\mathbf{w}(\omega)$ contains the complex-valued filter coefficients (for each frequency ω) and $\mathbf{d}(\omega, \Theta)$ denotes the commensurate steering vector representing the frequency- and direction-dependent transfer functions between the source and the N microphones. In order to calculate the filter coefficients $\mathbf{w}(\omega)$ one may employ a least-squares cost function J_{LS} , minimizing the sum over P directions of the squared absolute difference between $H(\omega, \Theta)$ and $D(\omega, \Theta)$, i.e.

$$J_{LS}(\mathbf{w}(\omega)) = \sum_{\Theta=1}^P |H(\omega, \Theta) - D(\omega, \Theta)|^2 . \quad (2)$$

The solution of Eq. 2 can be found in Rasumow *et al.* (2011) or Doclo and Moonen (2003). A straightforward minimization of Eq. 2, however, would result in non robust filter coefficients $\mathbf{w}(\omega)$, where already small errors of the microphone positions and/or characteristics may result in huge errors of the re-synthesized directivity patterns (cf. Rasumow *et al.* (2011)). This can be avoided by imposing a constraint on the so-called white noise gain (WNG) of the derived filter coefficients, defined as

$$\text{WNG}(\omega) = 10 \cdot \log_{10} \left(\frac{\mathbf{w}^H(\omega) \mathbf{w}(\omega)}{|\mathbf{w}^H(\omega) \mathbf{d}_{ref}(\omega)|^2} \right) , \quad (3)$$

with $\mathbf{d}_{ref}(\omega)$ being the steering vector from the frontal direction. In order to limit the WNG we suggest to apply the constraint $\text{WNG}(\omega) \leq \beta$. β has to be chosen manually according to the expected error of the steering vector (cf. Rasumow *et al.* (2011)). In the following x^H denotes the Hermitian transpose of x and x^* denotes the complex conjugate of x . The combination of the least squares cost function from Eq. 2 with the constraint from Eq. 3 results in the cost function

$$J_{LS\rho}(\omega) = \mu \cdot \underbrace{\left(\mathbf{w}^H(\omega) \mathbf{w}(\omega) - \beta_{pow} |\mathbf{w}^H(\omega) \mathbf{d}_{ref}(\omega)|^2 \right)}_{J_{reg}} + \underbrace{\sum_{\Theta=1}^P |\mathbf{w}^H(\omega) \mathbf{d}(\omega, \Theta) - D(\omega, \Theta)|^2}_{J_{LS}} , \quad (4)$$

where¹ μ represents the Lagrange multiplier. The closed form solution of $J_{LS\rho}(\omega)$, yielding the regularized filter coefficients $\mathbf{w}_{LS\rho}(\omega)$, is given by

$$\mathbf{w}_{LS\rho}(\omega) = \left(\mathbf{Q}(\omega) + \mu \left(\mathbf{I}_N - \beta_{\text{pow}} \cdot \mathbf{d}_{ref}(\omega) \mathbf{d}_{ref}^H(\omega) \right) \right)^{-1} \cdot \mathbf{a}(\omega) \quad \text{with,} \quad (5)$$

$$\mathbf{Q}(\omega) = \sum_{\Theta=1}^P \mathbf{d}(\omega, \Theta) \mathbf{d}^H(\omega, \Theta), \quad \mathbf{a}(\omega) = \sum_{\Theta=1}^P \mathbf{d}(\omega, \Theta) D^*(\omega, \Theta) \quad \text{and} \quad (6)$$

with \mathbf{I}_N the $N \times N$ -dimensional unity matrix. The optimization of μ as to satisfy the constraint $WNG(\omega) \leq \beta$ is described in section 6.

4. NON-LINEAR COST FUNCTION

From a psychoacoustical point of view, the least-squares cost function seems to be suboptimal, since a subjective sensation is approximately proportional to the logarithm of the stimulus intensity (cf. (Moore, 2003, Chapter 4)). Hence we propose to minimize the difference between the logarithm of the desired directivity pattern $D(\omega, \Theta)$ and the logarithm of the re-synthesized directivity pattern $H(\omega, \Theta)$ or its regularized counterpart, i.e.

$$J_{NL\rho}(\omega) = \underbrace{\mu \cdot \left(\frac{\mathbf{w}^H(\omega) \mathbf{w}(\omega)}{|\mathbf{w}^H(\omega) \mathbf{d}_{ref}(\omega)|^2} - \beta_{\text{pow}} \right)}_{J_{\text{reg}}} + \underbrace{\sum_{\Theta=1}^P \left| \log_{10}(\mathbf{w}^H(\omega) \mathbf{d}(\omega, \Theta)) - \log_{10}(D(\omega, \Theta)) \right|^2}_{J_{\text{NL}}}. \quad (7)$$

Note that both arguments ($\mathbf{w}^H(\omega) \mathbf{d}(\omega, \Theta)$ and $D(\omega, \Theta)$), as well as their difference, are complex-valued. By minimizing the squared magnitude of the complex difference, its real part as well as the imaginary part are taken into account. $J_{NL\rho}$ can be minimized, for instance, using various gradient descent methods. In order to enable the use of a trust-region method for the optimization, the analytical gradient of $J_{NL\rho}$ with respect to \mathbf{w} needs to be provided, given by

$$\begin{aligned} \frac{\partial J_{NL\rho}(\omega)}{\partial \mathbf{w}(\omega)} &= \frac{\partial J_{\text{reg}}(\omega)}{\partial \mathbf{w}(\omega)} + \frac{\partial J_{\text{NL}}(\omega)}{\partial \mathbf{w}(\omega)}, \quad \text{with} \quad (8) \\ \frac{\partial J_{\text{reg}}(\omega)}{\partial \mathbf{w}(\omega)} &= 2 \cdot \frac{\mathbf{w}(\omega) \mathbf{w}^H(\omega) \mathbf{d}_{ref}(\omega) \mathbf{d}_{ref}^H(\omega) \mathbf{w}(\omega) - \mathbf{w}^H(\omega) \mathbf{w}(\omega) \mathbf{d}_{ref}(\omega) \mathbf{d}_{ref}^H(\omega) \mathbf{w}(\omega)}{(\mathbf{w}^H(\omega) \mathbf{d}_{ref}(\omega) \mathbf{d}_{ref}^H(\omega) \mathbf{w}(\omega))^2} \quad \text{and} \\ \frac{\partial J_{\text{NL}}(\omega)}{\partial \mathbf{w}(\omega)} &= \sum_{\Theta=1}^P \left[\frac{\mathbf{d}(\omega, \Theta)}{(\mathbf{w}^H(\omega) \mathbf{d}(\omega, \Theta) \cdot \ln(10))} \cdot \left(\log_{10}(\mathbf{w}^H(\omega) \mathbf{d}(\omega, \Theta)) - \log_{10}(D(\omega, \Theta)) \right)^* \right. \\ &\quad \left. + \left(\frac{\mathbf{d}(\omega, \Theta)}{\mathbf{w}^H(\omega) \mathbf{d}(\omega, \Theta) \cdot \ln(10)} \right)^* \cdot \left(\log_{10}(\mathbf{w}^H(\omega) \mathbf{d}(\omega, \Theta)) - \log_{10}(D(\omega, \Theta)) \right) \right]. \end{aligned}$$

Analogous to the least-squares solution in Eq. 5, μ determines the impact of the regularization. The optimization of μ in order to satisfy the WNG constraint is described in section 6.

5. MERGED COST FUNCTION $J_{MIX\rho}(\omega)$

Within the application of the VAH it is essential to investigate how the subjective sound perception benefits from J_{NL} compared with J_{LS} . In order to do so, we introduce a merged cost

¹Note that when applying the desired WNG β in the optimization procedure, β must be provided as the power constant β_{pow} , with $\beta_{\text{pow}} = 10^{\frac{\beta}{10}}$ and $\beta = 10 \cdot \log_{10}(\beta_{\text{pow}})$.

function J_{MIX_ρ} , where the blending factor α determines the weight of J_{LS} and J_{NL} within the optimization of the filter coefficients $\mathbf{w}_{\text{MIX}_\rho}(\omega)$ by

$$J_{\text{MIX}_\rho}(\omega) = \underbrace{\mu \cdot \left(\frac{\mathbf{w}^{\text{H}}(\omega)\mathbf{w}(\omega)}{|\mathbf{w}^{\text{H}}(\omega)\mathbf{d}_{\text{ref}}(\omega)|^2} - \beta_{\text{pow}} \right)}_{J_{\text{reg}}} + \alpha \cdot \sum_{\Theta=1}^P \underbrace{\left| \log_{10}(\mathbf{w}^{\text{H}}(\omega)\mathbf{d}(\omega, \Theta)) - \log_{10}(D(\omega, \Theta)) \right|^2}_{J_{\text{NL}}} + (1 - \alpha) \cdot \sum_{\Theta=1}^P \underbrace{\left| \mathbf{w}^{\text{H}}(\omega)\mathbf{d}(\omega, \Theta) - D(\omega, \Theta) \right|^2}_{J_{\text{LS}}}. \quad (9)$$

The aim of this study was to evaluate the subjective perception using filter coefficients $\mathbf{w}_{\text{MIX}_\rho}(\omega)$, optimizing J_{MIX_ρ} with varying α . $J_{\text{NL}_\rho}(\omega)$ and consequently $J_{\text{MIX}_\rho}(\omega)$ cannot be solved in a closed form manner. Therefore we employed a gradient descent method to solve $J_{\text{MIX}_\rho}(\omega)$. In particular we used the function **fminunc** in MATLAB (Optimization Toolbox). In general when applying gradient descent methods such as **fminunc**, initial values must be provided. We used the filter coefficients $\mathbf{w}_{\text{LS}_\rho}$, derived from the closed form solution in Eq. 5, as initial values. Furthermore, in order to solve J_{MIX_ρ} one needs to provide an analytical expression for the gradient of J_{MIX_ρ} with respect to the filter coefficients \mathbf{w} , i.e.

$$\frac{\partial J_{\text{MIX}_\rho}(\omega)}{\partial \mathbf{w}(\omega)} = \frac{\partial J_{\text{reg}}(\omega)}{\partial \mathbf{w}(\omega)} + \alpha \cdot \frac{\partial J_{\text{NL}}(\omega)}{\partial \mathbf{w}(\omega)} + (1 - \alpha) \cdot \sum_{\Theta=1}^P \mathbf{d}(\omega, \Theta) \cdot (\mathbf{w}^{\text{H}}\mathbf{d}(\omega, \Theta) - D(\omega, \Theta))^* + \left((\mathbf{w}^{\text{H}}\mathbf{d}(\omega, \Theta) - D(\omega, \Theta)) \cdot \mathbf{d}^*(\omega, \Theta) \right). \quad (10)$$

6. FINDING A SUITABLE LAGRANGE MULTIPLIER μ FOR REGULARIZATION

As can be seen in Eq. 4, 7 and 9 the regularization parameter μ determines the weight of the regularization term J_{reg} within the cost function. In general, when taking a larger value for μ , the WNG approaches the desired value β while reciprocally the intrinsic term (J_{LS} or/and J_{NL}) is increased. Due to varying costs with a variable α , a constant μ does not guarantee a constant WNG. In order to ensure a comparable regularization between the different cost functions (and for varying α) we determined the regularization parameter μ such that the WNG is smaller than -3 dB. This constraint ensures that the inner product of the obtained filter coefficients $\mathbf{w}(\omega)$ is 3 dB smaller than the power of the filtered signal arriving from the front. Given the desired WNG β , the filter coefficients can be computed using an iterative optimization method. In this process one needs to find a suitable regularization parameter μ , resulting in $\text{WNG}(\mu) \leq -3$ dB for various α and individual HRTFs.

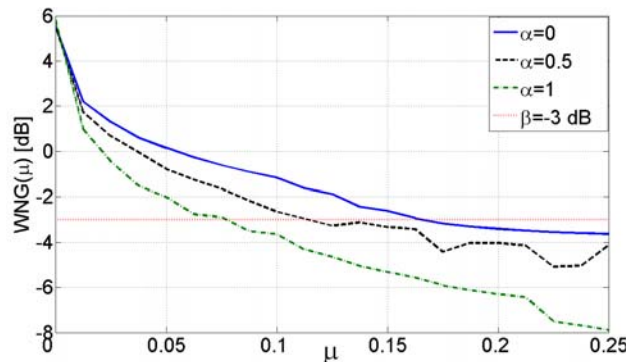


FIGURE 2: Exemplary course of the WNG as a function of μ optimizing J_{MIX_ρ} with three different α and $\beta = -3$ dB.

In general the WNG is assumed to decrease with larger μ up to a certain point. However, due to the use of a non-linear cost function, $WNG(\mu)$ is not necessarily a monotonous function. An exemplary course of the WNG as a function of μ with three different α is depicted in Fig. 2. Here it becomes apparent that $WNG(\mu)$ varies with the cost function, i.e. with α . We found that for the given population and $0 \leq \alpha \leq 1$ the desired WNG of -3 dB can typically be obtained with $\mu \leq 0.3$. Thus, in order to find the smallest μ resulting in the desired WNG we computed the filter coefficients $\mathbf{w}_{MIX_p}(\mu)$ for several ascending μ . More specifically, starting with $\mu = 0$ we computed $\mathbf{w}_{MIX_p}(\mu)$ and increased μ until $WNG(\mathbf{w}_{MIX_p}(\mu)) \leq -3$ dB was reached. In doing so, μ was increased incrementally in steps of $\Delta \mu = 0.02$ until the desired WNG or the upper limit of $\mu = 0.3$ was reached (if existent at all, this only occurred at very low frequencies). Using this brute-force and rather time-consuming method, the regularization parameter for each cost function can be optimized such that the desired WNG is obtained.

7. EXPERIMENTAL PROCEDURE

Prior to the experiment, the HRTFs and headphone (AKG K-240 Studio) transfer function (HPTF) of four subjects were measured individually. To this end the subjects were seated in an anechoic chamber, surrounded by 24 circularly arranged loudspeakers at a radius of 1.25 m. HRTFs and HPTFs were measured according to the blocked ear method (cf. Hammershoi and Moller (1996)) using individual custom-made ear shells and Knowles FG-23329 miniature electret microphones. The individual HRTFs as well as the inverse HPTFs were implemented as FIR-filters with a length of 512 taps, corresponding to ≈ 11.6 ms at a sampling frequency of $f_s = 44100$ Hz. As to cover a wide frequency range and simultaneously to include temporal cues, the test signal consisted of short bursts of white noise with a spectral content of $150 \text{ Hz} < f < 18050 \text{ Hz}$. The spectral amplitude of the noise bursts was faded in between 150 and 200 Hz and faded out between 18000 and 18050 Hz using tapered Hann windows to avoid a pitch cue due to sharp edges in the frequency domain. Each noise burst lasted 0.15 seconds with 0.01 seconds onset-offset ramps followed by silence of 0.15 seconds.

In order to investigate the perceptual effects introduced by fading between the aforementioned cost functions, we implemented an A/B comparison with the paradigm depicted in Fig. 3. Four subjects were instructed to compare the test signal filtered with the individual HRTFs (reference setting) and the test signal re-synthesized using the VAH (test setting) with a variable α . Within the reference setting the signal was binaurally filtered with corresponding HRTF filters and subsequently with the inverse HPTF filters. Currently this setting is supposed

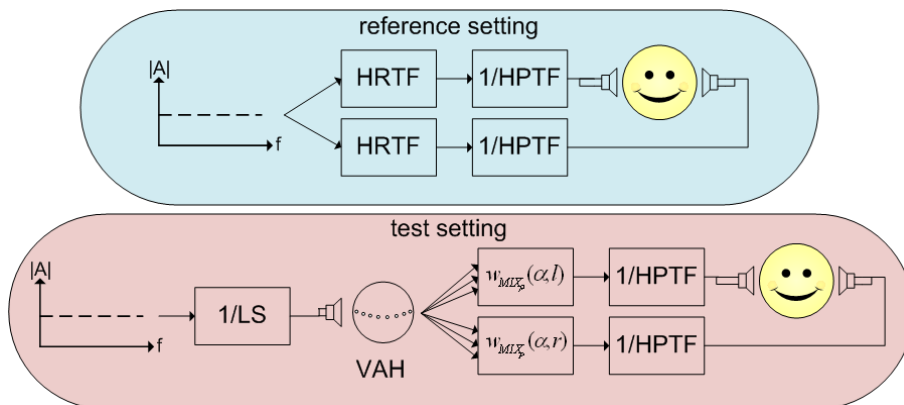


FIGURE 3: Testing paradigm of the implemented A/B comparison. Subjects were instructed to compare the signal from the reference and the test setting. The test setting was varied in randomized order by varying α ($\alpha = 0, \frac{1}{3}, \frac{2}{3},$ or 1), the azimuthal direction ($\theta = 0^\circ, 90^\circ$ or 225°) and three iterations for each condition.

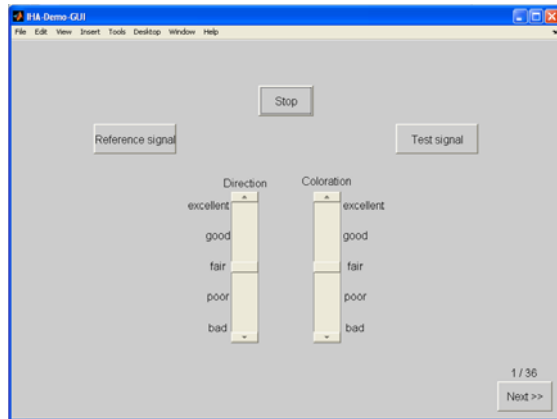


FIGURE 4: Graphical user interface (GUI). Subjects were instructed to evaluate the subjective difference with respect to the perceived direction (left slider) and spectral coloration (right slider).



FIGURE 5: VAH with 24 equidistantly spaced microphones (Sennheiser KE-4) on the horizontal plane. The microphones are mounted flush in a rigid sphere with a diameter of ≈ 18 cm.

to yield the best immersive reproduction of spatial hearing aspects (cf. Huopaniemi *et al.* (1999), Rasumow *et al.* (2012)). Thus the various VAH re-syntheses were evaluated in comparison to this reference setting within the following A/B comparison. Prior to the listening tests, the signal was recorded from the three azimuthal directions using the VAH in an anechoic chamber. Also the loudspeakers used for the playback were equalized individually (cf. 1/LS in the test setting in Fig. 3). The VAH consisted of 24 equidistantly spaced microphones on the equator of a rigid sphere with a diameter ≈ 18 cm (cf. Fig. 5). In each condition the 24 raw signals were filtered with corresponding filters $\mathbf{w}_{\text{MIX}_p}(\omega, \alpha)$ for each ear. The sum of the filtered signals, representing the re-synthesized ear signals, was filtered with the inverse HPTF filters and played to the subject via headphones.

To limit the number of experiments to a manageable amount, three directions in the horizontal plane were chosen with azimuth angles $\theta = 0^\circ$ (front), 90° (left) and 225° (back right) and α was one of $\alpha = 0, \frac{1}{3}, \frac{2}{3}$, and 1. The azimuthal direction θ as well as α were varied in randomized order within one experimental run with three iterations for each condition. This resulted in 36 conditions for each subject to be evaluated. For each condition the subject could switch between the reference and the test setting arbitrarily (cf. Fig. 4). The subjects were instructed to evaluate the difference between the reference setting and the test setting with reference to the perceived direction and spectral coloration. A category scale, ranging from 1 (bad) to 5 (excellent) in steps of 0.2 was used to rate the individual perception (cf. Fig. 4). The test signal was played back continuously with a cross-fading between the test and reference setting. Playback was stopped when the subject clicked on the *Next* button and the rating was saved. Furthermore, playback also could be stopped with an additional *Stop* button. Each session lasted approximately 30 minutes and subjects had time for familiarization before the test started.

8. RESULTS AND DISCUSSION

The results from the subjective evaluation with respect to direction and spectral coloration are depicted in Fig. 6 and Fig. 7 respectively. Here the mean and standard deviation (over three iterations) are plotted against α . The three azimuthal directions are separated by boxes and subjects are characterized by various colors/symbols.

In general the obtained results reveal fair to excellent evaluations (except for SK with regard to coloration at $\theta = 0^\circ$). In the main these results emphasize the subjective validity of the

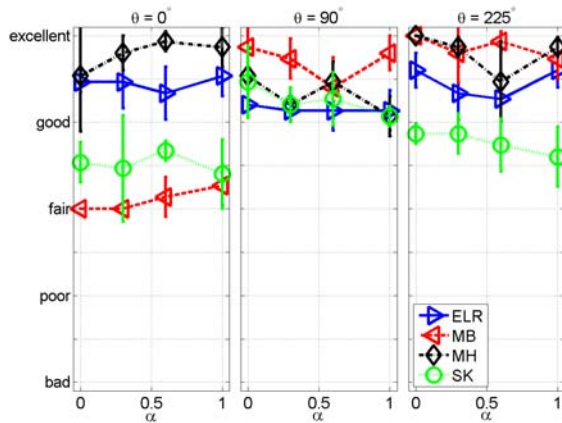


FIGURE 6: Subjective evaluation with respect to direction. The mean (over three iterations) and standard deviation for the evaluation of each condition is depicted as a function of α (x-axis) for three azimuthal directions. The colors and symbols characterize the four subjects.

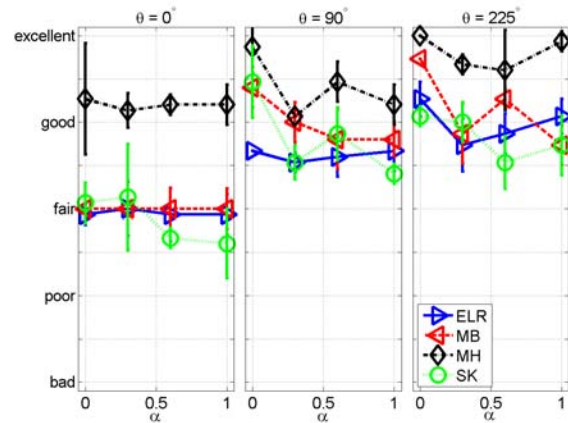


FIGURE 7: Subjective evaluation with respect to spectral coloration. The mean (over three iterations) and standard deviation for the evaluation of each condition is depicted as a function of α (x-axis) for three azimuthal directions. The colors and symbols characterize the four subjects.

concept of re-synthesizing HRTF with the VAH. Moreover, the evaluation is perspicuously best for the lateral directions with $\theta = 90^\circ$ & $\theta = 225^\circ$ while the re-synthesis for $\theta = 0^\circ$ is mostly associated with the largest difference compared to the reference setting and consequently with the worst evaluation. This phenomenon is prominently apparent in the evaluation with respect to spectral coloration (cf. Fig. 7), especially for subject ELR, MB and SK. Since all re-synthesized signals are based on previous recordings with the VAH it is possible that its positioning was slightly shifted for the test setting with $\theta = 0^\circ$. Such a positioning error of the VAH would result in slightly differing re-synthesized HRTFs and thus could explain the lower rating for $\theta = 0^\circ$ compared with $\theta = 90^\circ$ and $\theta = 225^\circ$. The obtained standard deviation σ approximately varies between $0 \leq \sigma \leq 1$ and does not seem to depend on the direction θ nor on the blending factor α .

Assuming that the re-synthesis using $J_{NL\rho}$ would lead to better evaluations compared with the re-synthesis using $J_{LS\rho}$ one would expect the rating (for a fixed condition) to increase with α in Fig. 6 and Fig. 7. At least one would assume the evaluation for $\alpha = 1$ to be significantly better than for $\alpha = 0$. However, this assumption does not correspond to the observed evaluation. In detail the results show trends for individual preferences. For instance, the results for subject MH do not seem to depend considerably on the tested directions while there is a distinct preference for the other three subjects. Yet, in consideration of the standard deviation none of the subjects tends to adjust the evaluation with respect to α . Especially the evaluation with respect to direction does not seem to depend unambiguously on α .

Interestingly, for some conditions the evaluation seems to decrease faintly with α . For instance, the evaluation with respect to coloration seems better for smaller α (subject MB, SK and $\theta = 90^\circ$ & 225°). On the other hand, the variation of the evaluation as a function of α has approximately the magnitude of the standard deviation arising from subsequent iterations, putting this vague tendency into perspective. The assumed improvement of the evaluation with increasing α is only observed vaguely for two conditions (subject MH and MB with $\theta = 0^\circ$ in Fig. 6). Yet, also this tendency is very weak and does not exceed the magnitude of the obtained standard deviation.

In sum the observed evaluation shows a general subjective acceptance for the re-synthesis of HRTFs using a VAH compared with measured HRTFs. However, there seems to be a directional preference which is assumed to be due to a slightly incorrect positioning of the VAH. The obtained evaluations do not confirm the hypothesis of enhancement by fading from a least-squares cost function $J_{LS\rho}$ to a non-linear cost function $J_{NL\rho}$. In consideration of the small

population (four subjects) and the standard deviation the obtained evaluation shows no clear tendency with respect to the applied cost functions.

9. CONCLUSIONS AND FURTHER WORK

The described findings confirm the validity of re-synthesizing HRTFs using a VAH in conjunction with a monaural cost function. Moreover, the subjective rating of the re-synthesis does not seem to increase with a non-linear cost function. Thus, for the application of a VAH, the derivation of filter coefficients using a least-squares cost function, as in Eq. 5, is preferred considering the advantages associated with a closed form solution.

The evaluation from Fig. 6 and Fig. 7 illustrates that there is still room for improvement in the derivation of filter coefficients. Thus, based on the presented framework, we will implement and evaluate analogous binaural cost functions in order to assess associated assets and drawbacks.

ACKNOWLEDGMENTS

This project was partially funded by Bundesministerium für Bildung und Forschung under grant no. 17080X10.

REFERENCES

- Atkins, J. (2011). "Robust beamforming and steering of arbitrary beam patterns using spherical arrays", in *Applications of Signal Processing to Audio and Acoustics (WASPAA), 2011 IEEE Workshop on*, 237–240.
- Doclo, S. and Moonen, M. (2003). "Design of broadband beamformers robust against gain and phase errors in the microphone array characteristics", *IEEE TRANSACTIONS ON SIGNAL PROCESSING* **51**, 2511–2526.
- Hammershoi, D. and Moller, H. (1996). "Sound transmission to and within the human ear canal", *The Journal of the Acoustical Society of America* **100**, 408–427.
- Huopaniemi, J., Zacharov, N., and Karjalainen, M. (1999). "Objective and subjective evaluation of head-related transfer function filter design", *J. Audio Eng. Soc* **47**, 218–239.
- Kahana, Y., Nelson, P. A., Kirkeby, O., and Hamada, H. (1999). "A multiple microphone recording technique for the generation of virtual acoustic images", *The Journal of the Acoustical Society of America* **105**, 1503–1516.
- Moore, B. C. J. (2003). *An Introduction to the Psychology of Hearing, Fifth Edition* (Academic Press).
- Rasumow, E., Blau, M., Hansen, M., Doclo, S., van de Par, S., Mellert, V., and Püschel, D. (2011). "Robustness of virtual artificial head topologies with respect to microphone positioning errors", in *Proc. Forum Acusticum, Aalborg*, 2251–2256 (Aalborg).
- Rasumow, E., Blau, M., Hansen, M., Doclo, S., van de Par, S., Püschel, D., and Mellert, V. (2012). "Smoothing head-related transfer functions for a virtual artificial head", in *Acoustics 2012*, 1019–1024 (Nantes, France).
- Tohtuyeva, N. and Mellert, V. (1999). "Approximation of dummy-head recording technique by a multimicrophone arrangement", *The Journal of the Acoustical Society of America* **105**, 1101–1101.