

# Examining the Influence of Controlled HRTF-Variation on Various Distance Metrics

Cosima A. Ermert, Shaima'a Doma, Janina Fels

*Institute for Hearing Technology and Acoustics, RWTH Aachen University, 52074 Aachen, Germany*

*Email: {cosima.ermert; shaimaa.doma; janina.fels} @akustik.rwth-aachen.de*

## Introduction

There are multiple possible reasons for comparing head-related transfer functions (HRTFs), e.g., to evaluate different measurement setups, to examine differences between simulated and measured HRTFs, for HRTF clustering, or to investigate various influences like wearing headgear or age-related changes (e.g., [1, 2, 3, 4]). However, comparing HRTFs is an intricate task due to their high data complexity resulting from information for different directions, frequencies and ears. Distance metrics offer specific computation algorithms to highlight differences between HRTFs.

This work aims to examine the behavior of several distance metrics based on controlled HRTF variation. For this analysis, the following distance metrics were chosen: mean-squared error (MSE) [3], critical bands mean-squared error (CB-MSE) [3], inter-subject spectral difference (ISSD) [2], spherical difference (SD) [1], mel-frequency cepstral distortion (MFCD) [5], and loudness level spectrum error (LLSE) [6]. First, the definitions of the distance metrics are introduced, followed by the procedure for controlled HRTF variation. Subsequently, results regarding the range of value, spatial distribution and influence of relative spectral shifting are presented and discussed.

## Definition of Distance Metrics

This paper focuses on a selection of six distance metrics, of which three distance metrics are purely based on numerical methods and three incorporate the frequency selectivity of the human auditory system on the basis of different psychoacoustic models. All distance metrics provide an error value in the directional domain, except for the SD, which gives insights on dissimilarities for each frequency bin.

To ease the comparison, sometimes a single value is desired instead of a directional data set or frequency spectrum. Usually, this is achieved by averaging over the directions or frequencies (e.g., [2]). However, when averaging over the sphere the HRTF spans over, the spatial sampling has to be considered [1]. In this paper, the HRTFs are sampled equiangularly. When averaging over direction-dependent metrics to get a single valued metric (indicated by  $\bar{d}$ ), this equiangular weighting is considered by employing corresponding voronoi weighting. No spectral weighting is applied in the single valued version of the frequency-dependent metric SD.

### Mean-Squared Error (MSE)

A well-known numerical method for comparing digital filters in signal processing is the mean-squared error

(MSE) criterion, which can also be used for comparing HRTFs [3]. The MSE is calculated for each direction  $k$  by subtracting the magnitude spectra  $H_1$  and  $H_2$  of the HRTF-set for each frequency bin, as follows:

$$d_{\text{MSE}}(k) = \frac{1}{M} \sum_{m=1}^M [H_1(f_m, \theta_k, \varphi_k) - H_2(f_m, \theta_k, \varphi_k)]^2, \quad (1)$$

where the index  $m$  represents the frequency index and  $M$  the total number of FFT-points. The direction  $k$  is specified by the corresponding azimuth angle  $\varphi_k$  and the elevation angle  $\theta_k$ .

### Critical Bands Mean-Squared Error (CB-MSE)

In [3], a variation of the MSE is presented, which incorporates a weighting factor  $\alpha(m)$  based on critical bands [7, p. 149 ff.]. Analogous to the auditory frequency resolution, which is better for low and poorer for high frequencies, the weighting factor decreases for increasing frequencies.

This critical bands mean-squared error (CB-MSE) is computed similarly to the MSE:

$$d_{\text{CB-MSE}}(k) = \frac{1}{M} \sum_{m=1}^M (\alpha(m)[H_1(f_m, \theta_k, \varphi_k) - H_2(f_m, \theta_k, \varphi_k)])^2. \quad (2)$$

### Inter-Subject Spectral Difference (ISSD)

The second numerical method in this paper is the inter-subject spectral difference (ISSD) [2]. Here, the focus is set on differences in the HRTF's fine structure. Therefore, the corresponding direction-dependent transfer functions (DTFs) are used instead of the HRTF set. The DTF is a variation of the HRTF, which omits information common to all directions. The variance of the DTF ratio is computed over frequencies:

$$d_{\text{ISSD}}(k) = \sigma^2 \left( 20 \log_{10} \frac{|\text{DTF}_1(f_m, \theta_k, \varphi_k)|}{|\text{DTF}_2(f_m, \theta_k, \varphi_k)|} \right). \quad (3)$$

### Spherical Difference (SD)

A variation of the ISSD, which yields frequency- instead of direction-dependent information, is the spherical difference (SD) [1]. The three main changes compared to the ISSD are the use of the standard deviation instead of the variance, computing it over directions instead of frequencies, and introducing a direction-dependent weighting factor to account for the spatial sampling.

$$d_{\text{SD}}(m) = \sigma_w \left( 20 \log_{10} \frac{|\text{DTF}_1(f_m, \theta_k, \varphi_k)|}{|\text{DTF}_2(f_m, \theta_k, \varphi_k)|} \right) \quad (4)$$

### Mel-Frequency Cepstral Distortion (MFCD)

Mel-frequency cepstral coefficients (MFCCs) are often used in speech processing applications. They provide a compact representation of the frequency spectrum of a signal by exploiting the subjectively perceived pitch. MFCC-spectra can also be used to compare HRTFs [5]. The resulting distance metric, the mel-frequency cepstral distortion (MFCD) is calculated as:

$$d_{\text{MFCD}}(k) = \frac{1}{N_C} \sum_{n=1}^{N_C} (M_1(n, k) - M_2(n, k))^2, \quad (5)$$

where  $M_1(n, k)$  and  $M_2(n, k)$  are the  $n$ th MFCC for the direction  $k$ .  $N_C$  is the total number of MFCCs, here 24. For a detailed description on how to compute the MFCCs, please refer to [8].

### Loudness Level Spectrum Error (LLSE)

The loudness level spectrum error (LLSE) is based on an algorithm presented in [6]. First, a pink noise signal is filtered with  $\text{HRTF}_1$  and  $\text{HRTF}_2$  respectively. In order to model the frequency selectivity of the human auditory system, the filtered noise is passed through a gammatone filter-bank with  $B = 42$  channels [9]. For each channel  $b$ , the resulting loudness level  $L_L$  is determined. To capture differences in coloration over frequency, the variation of the loudness level difference is computed:

$$d_{\text{LLSE}}(k) = \sigma^2(L_{L1}(b, k) - L_{L2}(b, k)). \quad (6)$$

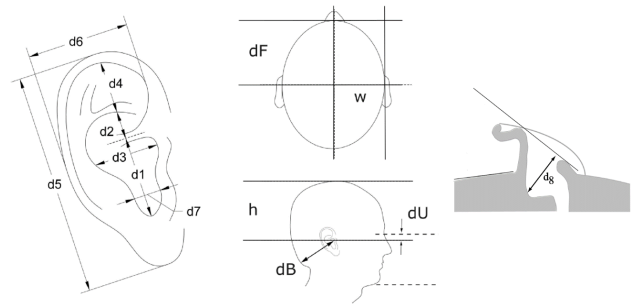
### Controlled HRTF Variation

To be able to compare the distance metrics in a meaningful way, a database consisting of HRTF which are varied stepwise in a controlled manner is needed. These requirements are hard to meet with measured HRTFs as they are error-prone, e.g., due to subject movement, and variation possibilities are limited in terms of control. Consequently, a database of simulated HRTFs was used.

The HRTFs were simulated using multiple regression analysis (MRA) and principal component analysis (PCA). PCA allows for a compressed representation of multivariate data. For that purpose, the data is represented by linear combinations of principal components (PCs), benefiting of trends in variance shared by many instances, here HRTF frequency spectra. By applying this procedure to an HRTF database, it is possible to reconstruct HRTFs in the database using linear combinations of the PCs.

Furthermore, with MRA, the weighting factors of the linear combinations can be linked to certain anthropometric features of subjects in the database. Thus, also new HRTF sets for arbitrary listeners can be generated based on their known anthropometric dimensions [10, 11]. As a result, it is possible to vary HRTFs in a controlled manner by changing anthropometric dimensions stepwise.

In the present study, the MRA of PCs was performed with the aid of the IHTA-toolbox [13] in MATLAB using the IHTA-database [12], which consists of 3D-scans and individually measured HRTF of 47 adults. The anthropometric dimensions considered for this procedure



**Figure 1:** Anthropometric dimensions in the IHTA-database (from [11]).

can be seen in Fig. 1. Using correlation analysis, the head height  $h$ , cymba concha height  $d_2$ , cavum concha width  $d_3$ , pinna width  $d_6$ , intertragal incisure width  $d_7$  and cavum concha depth  $d_8$  were identified to be the least correlated features. Thus, the stepwise variation was only based on these dimensions to avoid collinearity problems in the regression model [14]. Their statistical properties are displayed in Tab. 1. For this paper, a reference HRTFs was generated from average anthropometric dimensions ( $\mu$  values in Tab. 1). After [11], a parameter dependent step size of  $\Delta = 0.1 \cdot \sigma$  was chosen for the HRTF variation. Using a step number  $n$ , new HRTFs were generated on the basis of varying anthropometric dimensions after  $\mu + n \cdot \Delta$ .

**Table 1:** The six least correlated anthropometric dimensions in the IHTA database [12] for controlled HRTF variation: statistical properties are given by the mean  $\mu$  and the standard deviation  $\sigma$ .

	$\mu$ [mm]	$\sigma$ [mm]
Head height $h$	133.19	6.31
Cymba concha height $d_2$	8.68	1.98
Cavum concha width $d_3$	18.25	2.75
Pinna width $d_6$	28.34	3.10
Intertragal incisure width $d_7$	5.34	1.42
Cavum concha depth $d_8$	14.70	1.70

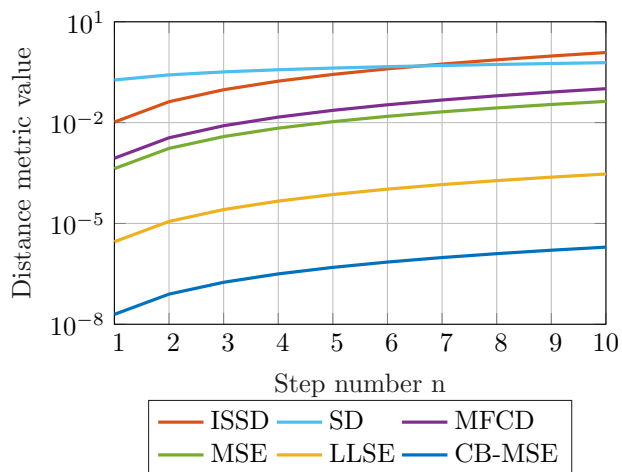
## Results and Discussion

In the scope of this paper, three research questions were followed:

- How can comparability in between metrics be ensured?
- How do the direction-dependent distance metrics behave over space?
- How do the distance metrics react to spectral misalignment?

### Comparability of Distance Metrics

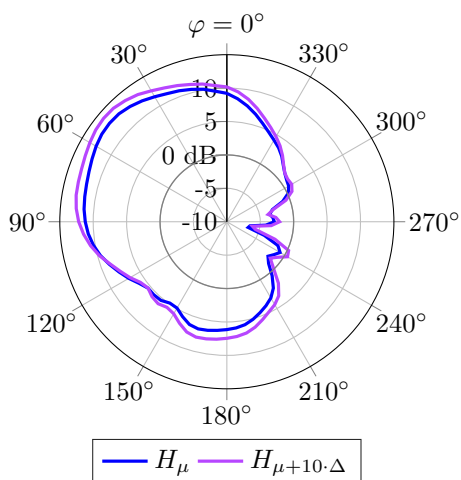
The range of values highly influences comparability between distance metrics. To evaluate the range of values, the distance for each metric was computed between  $H_\mu$  and the variations  $H_{\mu+n \cdot \Delta}$  for the step numbers  $n \in \{1, \dots, 10\}$ . The resulting values are depicted in Fig. 2. Although all distance metrics are very different



**Figure 2:** Distance metric values between  $H_\mu$  and the variations  $H_{\mu+n\Delta}$ , with  $\Delta = 0.1 \cdot \sigma$ , for the step number  $n$ .

in value range, they show a similar slope on the logarithmic axis, with the frequency-dependent SD being an outlier. The large value range complicates direct comparability between metrics.

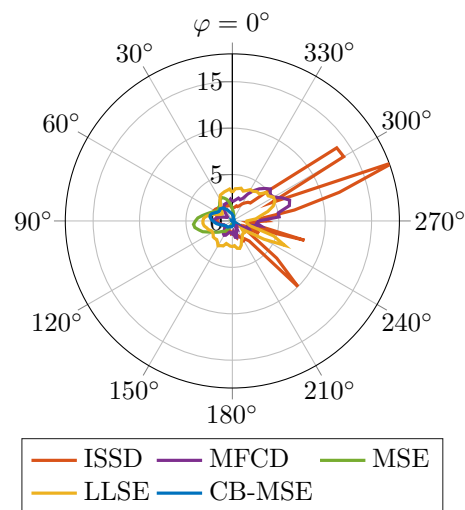
This issue could be resolved by introducing a metric dependent scaling factor, which transforms all metrics into the same value range. However, this is an intricate task. One possible approach could be to transform the values into a space between the minimum and maximum reachable value for each metric, denoted by 0 and, respectively, 1. This opens up the question of how the maximum reachable value, i.e., a set of two "most dissimilar" HRTF, can be achieved. In this paper, this maximum value is defined as the average value  $\bar{d}$  between  $H_\mu$  and  $H_{\mu+10\Delta}$  for each metric. Obviously, this is no perfect solution, since single direction- or frequency-dependent distance metric values can be higher than the weighted average  $\bar{d}$ . Still, this compromise ensures comparability in the scope of this paper.



**Figure 3:** Reference and varied HRTF sets  $H_\mu$  and  $H_{\mu+10\Delta}$  on the horizontal plane (elevation  $\theta = 0^\circ$ ) in dB at 3.5 kHz.

### Spatial Distribution of Distance Metrics

For this evaluation, only the direction-dependent distance metrics (MSE, CB-MSE, ISSD, MFCD, and LLSE)



**Figure 4:** Distance metric values for comparing  $H_\mu$  and  $H_{\mu+10\Delta}$  on the horizontal plane (elevation  $\theta = 0^\circ$ ), scaled with reference to  $\bar{d}$  between  $H_\mu$  and  $H_{\mu+10\Delta}$ .

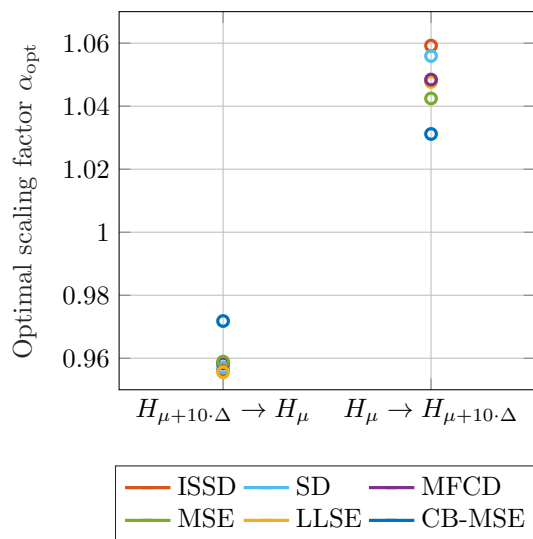
are considered. In Fig. 3, the spatial distribution of left-ear data sets  $H_\mu$  and  $H_{\mu+10\Delta}$  on the horizontal plane is displayed, exemplary at 3.5 kHz. The resulting distance values show maxima at different azimuth angles  $\varphi$  (see Fig. 4).

For the MSE and CB-MSE, which rely on absolute magnitude differences without further normalization, higher distance values are visible in the left hemisphere. Here, the HRTFs themselves have a higher magnitude, thus larger absolute differences are more probable. On the contrary, the MFCD and LLSE are more dominant in the right hemisphere, where "secondary" waves of small magnitude are revealed due to the shadowing effect of the head. Here, a detailed distribution of constructive and destructive interference can be observed at  $\varphi \approx 270^\circ$ , leading to a stronger variance in magnitude, both in space and in frequency [15]. Both MFCD and LLSE seem to give insights on the alignment of these structures. Lastly, the ISSD shows prominent peaks in the right hemisphere, namely at angles, where the larger HRTF  $H_{\mu+10\Delta}$  takes values smaller than  $H_\mu$ , e.g., at  $\varphi = 235^\circ$ . Since these notches further vary with elevation  $\theta$  and frequency, this might lead to higher variance in the DTF ratio (see Eq. 3) and thus in ISSD value.

### Spectral Alignment

In [2], frequency scaling was introduced as an HRTF individualization technique. The underlying idea is that an HRTF of a larger person can be adjusted to a smaller person by shifting the spectra towards higher frequencies, and vice versa. Successful individualization is achieved with the *optimal scaling factor*  $\alpha_{\text{opt}}$ , where the averaged ISSD between the shifted HRTF of another individual and the actual HRTF becomes minimal.

In the present study, applying this concept to other distance metrics showed that  $\alpha_{\text{opt}}$  yields comparable results throughout all considered metrics (see Fig. 5). Here,  $\alpha_{\text{opt}}$  was computed for shifting  $H_{\mu+10\Delta}$  towards  $H_\mu$  and vice versa.



**Figure 5:** Optimal scaling factor  $\alpha_{\text{opt}}$  for shifting  $H_{\mu+10\Delta}$  towards  $H_{\mu}$  ( $H_{\mu+10\Delta} \rightarrow H_{\mu}$ ), and vice versa.

This, however, only applies if the HRTFs are rather similar. With a step number  $n$  of 10, the absolute variation of each anthropometric dimension is still smaller than 6 mm. When comparing  $H_{\mu}$  with, e.g., an HRTF from maximum anthropometric features in the database, no unified  $\alpha_{\text{opt}}$  can be found for all distance metrics.

## Conclusion

In this paper, six distance metrics were presented. Using controlled HRTF variation, their behavior was analyzed. Regarding comparability, all distance metrics cover different ranges of values. Thus, a meaningful scaling factor is necessary to enable direct comparison between metrics. Furthermore, the slope of the frequency-dependent metric SD differed from the other metrics. This opens up the question whether this behavior is caused by the frequency dependency, meaning that other frequency-dependent metrics would behave similarly. Further investigations are necessary.

Analyzing the spatial distribution gave insight on which of the HRTF characteristics influences each distance metric the most. While MSE and CB-MSE are most sensitive regarding absolute magnitude, the ISSD yields high values when the HRTF spectra intersect. The MFC and LLSE showed higher distance values for the contralateral hemisphere, where the head shadowing effect causes a detailed distribution of constructive and destructive interference. These interference patterns are considered to be relevant for localization. Thus, it is possible that LLSE and MFC give insights on localization performance. The optimal scaling factor for frequency scaling after [2] is not identical but comparable between metrics for moderately varied HRTFs. To be able to compare HRTFs in a meaningful way, further investigations are needed to assess how distance metric values and subjective perception can be linked.

## Funding Acknowledgements

This work was funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation): FE 168/6-1, Individual Binaural Synthesis of Virtual Acoustic Scenes.

## References

- [1] Richter, J.; Behler, G.; Fels, J. (2016): Evaluation of a Fast HRTF Measurement System. Proc. 140th Audio Eng. Soc. Conv.
- [2] Middlebrooks, J. C. (1999): Individual differences in external-ear transfer functions reduced by scaling in frequency. JASA 106 (3), 1480-1492
- [3] Nicol, R.; Lemaire, V.; Bondu, A.; Busson, S. (2006): Looking for a Relevant Similarity Criterion for HRTF Clustering: A Comparative Study. Proc. 120th Audio Eng. Soc. Conv., 4
- [4] Pörschmann, C.; Arend, J. M.; Gillioz, R. (2019): How wearing headgear affects measured head-related transfer functions. Proc. EAA SASP, 49–54
- [5] Shimada, S.; Hayashi, N.; Hayashi, S. (1994): A Clustering Method for Sound Localization Transfer Functions. J. Audio Eng. Soc. 42 (7/8), 577-584
- [6] Huopaniemi, J.; Zacharov, N.; Karjalainen, M. (1998): Objective and Subjective Evaluation of Head-Related Transfer Function Filter Design. J. Audio
- [7] Zwicker, E.; Fastl, H. (1999): Psychoacoustics - Facts and Models. Second Updated Edition. Springer
- [8] Lee, K.; Lee, S. (2011): A Relevant Distance Criterion for Interpolation of Head-Related Transfer Functions. IEEE Trans. Audio Speech Lang. Process. 19 (6), 1780–1790 Eng. Soc. 47
- [9] Pulkki, V.; Karjalainen, M.; Huopaniemi, J. (1999): Analyzing Virtual Sound Source Attributes Using a Binaural Auditory Model. J. Audio Eng. Soc. 47 (4), 203-217
- [10] Bomhardt, R.; Braren, H.; Fels, J. (2017): Individualization of head-related transfer functions using principal component analysis and anthropometric dimensions. Proc. Mtgs. Acoust. 29
- [11] Doma, S.; Braren, H.; Fels, J. (2019): Kleinste wahrnehmbare Unterschiede in der HRTF-Rekonstruktion aus Haupt-komponenten auf Basis von anthropometrischen Abmessungen. Fortschritte der Akustik - DAGA 2019, 611-614
- [12] Bomhardt, R.; de la Fuente Klein, M.; Fels, J. (2016): A high-resolution head-related transfer function and three-dimensional ear model database. Proc. Mtgs. Acoust. 172
- [13] Berzborn, M.; Bomhardt, R.; Klein, J. Ch.; Richter, J.; Vorländer, M. (2017): The ITA-Toolbox: An Open Source MATLAB Toolbox for Acoustic Measurements and Signal Processing. Fortschritte der Akustik - DAGA 2017
- [14] Stewart, G. W. (1987): Collinearity and least squares regression. Statistical Science
- [15] Avendano, C.; Duda, R. O.; Algazi, V. R. (1999): Modeling the contralateral HRTF. J. Audio Eng. Soc.