DOI: 10.61782/fa.2025.0228

FORUM ACUSTICUM EURONOISE 2025

COMPARING MEASURES OF INFORMATION IN HEAD-RELATED TRANSFER-FUNCTIONS

Nils Meyer-Kahlen[1*] **Katharina Pollack**[2] **Pedro Lladó**[3]
[1] Department of Information and Communications Engineering, Aalto University, Espoo, Finland
[2] Acoustics Research Institute, Austrian Academy of Sciences, Vienna, Austria
[3] Institute of Sound Recording, University of Surrey, Guildford, United Kingdom

ABSTRACT

The NEMO initiative works towards agreeing on one set of head-related transfer functions (HRTFs) for use across various applications whenever incorporating individual(ized) HRTFs is not feasible. This initiative is grounded on the assumption that listeners benefit from adapting to a set of HRTFs that is different from their own. Naturally, one key step in this process is the selection of the particular set. We hypothesize that the available information in the set influences adaptation speed and/or post-adaptation localization performance and, therefore, is a potential factor in the selection. To test this hypothesis, it is essential to develop measures that quantify the information of an HRTF set. To this end, this contribution presents several potential measures of information content and compares them using multiple HRTF sets.

1. INTRODUCTION

Human listeners have the ability to adapt to a set of head-related transfer-functions (HRTFs) that is different from their own, see [1] for a review. The recently introduced NEMO initiative tries to leverage this finding by proposing a default set of HRTFs, which could be used across applications whenever adopting individual(ized) HRTFs is not feasible [2]. Naturally, selecting a particular set is a critical step in developing this initiative, for which several considerations need to be made. Some of them are of a practical nature, such as usability for developers and flexibility with regard to the available spatial grids and formats. Apart from these, however, it is desirable to select a set to which people can adapt the best, which could be measured by, e.g., speed of adaptation and/or post-adaptation localization performance. Determining what makes a set of HRTF *adaptable* in this sense requires new scientific insight.

We hypothesize that adaptability depends on the *information* contained in the HRTF set, and adaptation studies are being planned to determine whether this is the case. A natural prerequisite for such studies, however, is a measure that quantifies the information contained in a set of HRTFs. When first introducing the question of adaptability, we presented a simple, ad-hoc measure termed *discriminability* [2].

In this contribution, we compare the discriminability against two other possible measures. The first one is the spectral entropy, which is directly related to the effective rank of the HRTF matrix, introduced in [3]. The second measure is the information gain (or, equivalently, mutual information), which recently has been used to study how choosing different feature sets influences the information within one set of HRTFs [4]. The information gain arises naturally in a Bayesian inference view on human sound localization, which is currently becoming

***Corresponding author**: *nils.meyer-kahlen@aalto.fi.*

more common [5–9]. While the effective rank acts directly on the HRTF data, the discriminability and the information gain can be computed from the results of different auditory models of sound localization from the auditory modeling toolbox [10]. For the discriminability, we use `baumgartner2014` [11], as in [2]. For the information gain, we also included `barumerli2023` [7]. As the magnitude of the information gain values depends on the model parameters, we also studied their influence on the outcome.

In the next section, we introduce the measures we compare in detail. In Section 3 we show computational results for the HUTUBS set of HRTFs. Then, in Section 4 we compare the results and show relations between the measures, including an analysis of how the selection of the model parameters affect the results.

2. MEASURES

In the following section, we present three measures for quantifying information in a set of HRTFs: discriminability [2], spectral entropy, which is directly related to the effective rank [3], and information gain [4].

2.1 Discriminability

In [2], we proposed to find the most informative HRTF using a measure that was computed based on localization predictions of an auditory model. To define the measure, we denote the HRTF of one subject p from one direction $\boldsymbol{\theta}_i\ \forall i\ \in [1, I]$ as $\boldsymbol{h}_p(\boldsymbol{\theta}_i) \in \mathbb{C}^{2F}$, where F is the number of frequency bins. Therefore, the response of the left and the right ear are stacked underneath each other. A complete set with I directions can then be represented as a matrix $\boldsymbol{H}_p \in \mathbb{C}^{2F\times I}$.

We then defined the discriminability for one subject's HRTFs $\boldsymbol{H}_p$ as the average variance of the model predictions given a set of I directions as

$$D_p = \sqrt{\frac{1}{I}\sum_{i=1}^{I} \mathrm{Var}\{p(\hat{\boldsymbol{\theta}}|\boldsymbol{H}_p, \boldsymbol{\theta}_i)\}}, \qquad (1)$$

where $\hat{\boldsymbol{\theta}}$ is the predicted, perceived direction. The term $p(\hat{\boldsymbol{\theta}}|\boldsymbol{H}_p, \boldsymbol{\theta}_i)$ can be understood as the probability of localizing the sound at $\hat{\boldsymbol{\theta}}$, given a source at $\boldsymbol{\theta}_i$ and the specific set of HRTFs $\boldsymbol{H}_p$. The perceived direction, $\hat{\boldsymbol{\theta}}$, technically is continuous, but in practice, a grid of J possible answers $\hat{\boldsymbol{\theta}}_j$ is assumed, making p a probability mass function. In Bayesian terms, $p(\hat{\boldsymbol{\theta}}|\boldsymbol{H}_p, \boldsymbol{\theta}_i)$ is called the posterior.

For the data presented, we used directions $\boldsymbol{\theta}_i$ in the median plane and the `baumgartner2014` model proposed in [12] to obtain the posterior. This is obtained by setting both template and target HRTFs to $\boldsymbol{H}_p$.

The intuition behind this choice was that a wide posterior means less specific localizability. We then assumed that since template and target are set to the same set, a wide posterior can only originate from less useful information content of that particular set. The measure was computed for simulated HRTFs from the HUTUBS database [13] and the most informative set of HRTFs was determined, but no further connection to information theory was given, and the measure was not related to other measures from the literature.

2.2 Effective Rank

One of the few measures of HRTF information content proposed in previous literature is the effective rank of the HRTF matrix [3]. The effective rank is based on the singular values of the HRTF matrix of each participant, $\boldsymbol{H}_p$. It is defined as

$$\mathcal{R}(\boldsymbol{H}_p) = \exp\left(-\sum_{n=1}^{Q} \bar{\sigma}_n^2 \log \bar{\sigma}_n^2\right), \qquad (2)$$

where

$$\bar{\sigma}_n^2 = \frac{\sigma_n^2}{\sum_{k=1}^{Q} \sigma_k^2}, \qquad (3)$$

are the normalized singular values of $\boldsymbol{H}_p$. Interestingly, the argument of the exponential function in Eq. (2) is the (Shannon) entropy of the normalized singular values, which is then also called the spectral entropy (SE)

$$\mathrm{SE} = -\sum_{n=1}^{Q} \bar{\sigma}_n^2 \log \bar{\sigma}_n^2. \tag{4}$$

This already reminds of information theoretical measures. Note that this measure only depends on the HRTF data, not on any model of human localization. A detailed derivation of the effective rank from random processes is given in [14]. For comparison below, we use SE with a logarithm of base 2, as the result can then be interpreted as having the unit *bits*.

2.3 Information Gain

While the structure of the SE already hints at deeper roots in information theory, it is not derived from principled assumptions about human perception and cognition. To this end, the so-called information gain (IG) offers more insights. The term is used synonymously with the term "mutual information" [4]. We prefer the term information gain, as the measure literally quantifies how much directional information a subject can gain from listening to sound sources. So far, the quantity has been used for comparing how much information about a sound source location is gained from various sets of auditory cues [4]. In general, it is defined as

$$\mathrm{IG} = H(\hat{\boldsymbol{\theta}}) - H(\hat{\boldsymbol{\theta}}|\boldsymbol{x}), \tag{5}$$

which is the difference between the entropy of the prior distribution $p(\hat{\boldsymbol{\theta}})$, modeling the internal beliefs about where sound sources are located if no acoustic input were given,

$$H(\hat{\boldsymbol{\theta}}) = -\sum_{j=1}^{J} p(\hat{\boldsymbol{\theta}}_j) \log_2 p(\hat{\boldsymbol{\theta}}_j), \tag{6}$$

and the conditional entropy

$$H(\hat{\boldsymbol{\theta}}|\boldsymbol{x}) = -\sum_{j=1}^{J} p(\hat{\boldsymbol{\theta}}_j|\boldsymbol{x}) \log_2 p(\hat{\boldsymbol{\theta}}_j|\boldsymbol{x}) \tag{7}$$

of the posterior distribution $p(\hat{\boldsymbol{\theta}}|\boldsymbol{x})$ of the perceived source locations given the specific acoustic signal at the two ears $\boldsymbol{x} \in \mathbb{C}^{2F}$. If the logarithm with base 2 is used, the result has the unit of *bits*. Estimates of the posterior can be obtained from various auditory models, such as the Bayesian models presented in [5] and [15].

While Eq. (5) captures the general concept of the information gain for any acoustic input $\boldsymbol{x}$, we need to specify the choices for the IG computed for comparing sets of HRTFs more precisely. We assume localization of a single sound source, where the signal is simply an impulse, i.e., the HRTF itself, $\boldsymbol{x} = \boldsymbol{h}_p(\boldsymbol{\theta}_i)$. Moreover, we are interested in the information gain not for one source direction $\boldsymbol{\theta}_i$, but would like to include an entire distribution of source directions. This turns the information gain into the conditional information gain, which can be expressed as

$$\mathrm{IG}_p = \sum_i p(\boldsymbol{\theta}_i) \left(H(\hat{\boldsymbol{\theta}}|\boldsymbol{\theta}_i) - H(\hat{\boldsymbol{\theta}}|\boldsymbol{H}_p(\boldsymbol{\theta}_i), \boldsymbol{\theta}_i) \right) \tag{8}$$

where $p(\boldsymbol{\theta}_i)$ is the distribution over the tested directions. The result should reflect the information gain that a particular listener obtains from listening to sound sources from all directions, rather than simply picking directions based on previous knowledge, i.e., without any recent listening.

Note that $p(\boldsymbol{\theta})$ and $p(\hat{\boldsymbol{\theta}})$ are two different prior distributions. The former is the distribution of tested directions, which could be chosen based on a likely distribution of source directions encountered in the real world. On the other hand, $p(\hat{\boldsymbol{\theta}})$ is the prior distribution of the perceived sound source direction, i.e., the distribution of direction that a listener would pick without any acoustic information. Setting $p(\boldsymbol{\theta})$ and $p(\hat{\boldsymbol{\theta}})$ to the same distribution implies that, without meaningful acoustic information, perceived localization follows the prior distribution of sounds occurring in the world. This represents a reasonable assumption. However, as we have no reliable data on such a distribution, we use a uniform distribution for both, giving equal weight to all directions.

3. EXPERIMENTS

The measures are computed for the simulated HRTFs of 96 participants contained in the HUTUBS dataset [13]. D is computed as in [2]. SE was computed after re-sampling the measurement grid to a quasi-uniform grid of 1300 points following a t-design. IG was computed for two models, resulting in $\mathrm{IG}_{\mathrm{Barum}}$ for `barumerli2023` and $\mathrm{IG}_{\mathrm{Baum}}$ for `baumgartner2014`. For $\mathrm{IG}_{\mathrm{Barum}}$, the same t-design was used as for the SE.

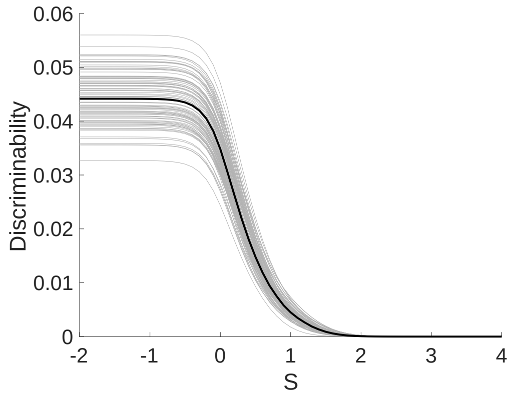
As the absolute values of the model-based measures depend on model parameters, different choices are studied for each of the models, before comparing the measures with each other using default parameters.

3.1 Discriminability

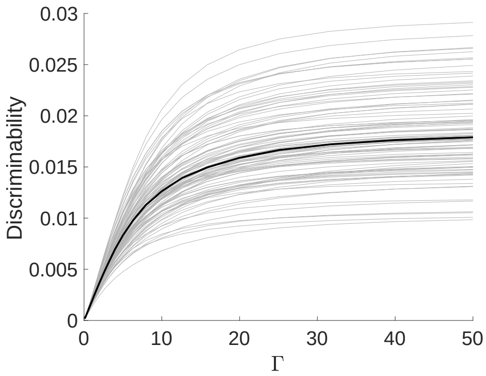
We computed the discriminability per set of HRTFs D by treating the probability mass vector obtained from `baumgartner2024` as posterior. The model has three fitted parameters: the degree of selectivity Γ, the listener sensitivity S and the sensorimotor scatter ε. In our experiments, the latter was set to $\varepsilon = 0$; this parameter was motivated by perception-to-pointing misalignments, which can be excluded when comparing sets of HRTFs.

The other two parameters have an influence on the discriminability which is shown in Fig. 1. It decreases as a function of S and increases as a function of Γ, which is expected from the parameter definition in [12] (note that lower S represents higher sensitivity). This implies that the discriminability increases for more sensitive and more selective parameterizations. These parameters are tuned either at group or subject level [12, 16] based on a specific localization task.

Importantly, Fig. 1 shows that the order between subjects (gray lines) is mostly maintained when varying the parameters, meaning that HRTFs can still be compared without dedicated tuning of the parameters. Since it is not obvious which parameters to use, they were set to the default values in [12], which are $\Gamma = 6\ \mathrm{dB}^{-1}$ and S = 1.



(a) `baumgartner2014` - S.



(b) `baumgartner2014` - Γ.

Figure 1: Influence of the parameters S and Γ from `baumgartner2014` on the Discriminability. Each gray line represents an individual subject from the database. The black lines represents the median over subjects.

3.2 Spectral Entropy

As mentioned before, the effective rank, and equivalently, the spectral entropy, does not depend on any model of human perception. It is merely a property of the HRTF matrix. Therefore, it does not depend on any model parameters either. Computing the spectral entropy resulted in values between $\mathrm{SE}^{\mathrm{min}} = 3.52$ bits and $\mathrm{SE}^{\mathrm{min}} = 4.28$ bits, with a median of $\mathrm{SE}^{\mathrm{median}} = 3.97$ bits.

3.3 Information gain

As with D, IG is computed after obtaining the posterior using auditory models. First, the IG was computed for `baumgartner2014` using the default parameterization of Γ and S ($\varepsilon = 0$, see the justification in Section 3.1). The results of computing the IG were relatively low ($\mathrm{IG}_{\mathrm{Baum}}^{\mathrm{median}}$ = 0.39 bits, $\mathrm{IG}_{\mathrm{Baum}}^{\mathrm{min}}$ = 0.17 bits, $\mathrm{IG}_{\mathrm{Baum}}^{\mathrm{max}}$ = 0.67 bits).

Figure 2 (a) and (b) show the effect of varying Γ while S is fixed and vice versa. The plots show that the absolute value of IG varies substantially depending on the specific parameters, by up to 3 bits. As with D, IG increases with higher sensitivity (lower S) and higher degree of selectivity (higher Γ). Again, the parameters seem to have a similar effect on all listeners, and they are not expected to have a significant impact on the correlation between different metrics if the same parameters are used for all sets of HRTFs included.

The IG was also computed after obtaining the posteriors returned by `barumerli2023`. They were computed using the default parameterization ($\sigma_{\mathrm{itd}} = 0.569$, $\sigma_{\mathrm{ild}} = 1$, $\sigma_{\mathrm{mon}} = 1.25$) and using a uniform prior distribution. In this model, the motor noise does not affect the posterior distribution and is used to add noise to the posterior-to-response mapping. To speed up the processing time, the number of repetitions of the experiment was set to only two. Systematic differences are discovered regardless.

The results with the default parameterization was $\mathrm{IG}_{\mathrm{Barum}}^{\mathrm{median}}$ = 7.20 bits, $\mathrm{IG}_{\mathrm{Barum}}^{\mathrm{min}}$ = 6.81 bits, $\mathrm{IG}_{\mathrm{Barum}}^{\mathrm{max}}$ = 7.49 bits. Varying one parameter while fixing the rest using only two runs of the model for speed shows that, as expected, increasing the noise for localization cues (σ_{itd}, σ_{ild}, σ_{mon}) decreased the IG, having σ_{mon} the largest impact, see Figure 2 (c), (d), (e) and (f). Moreover, the effect of varying the prior distribution was analyzed. The results show that informative priors (low σ_{prior}) decrease the IG that acoustic cues provide. This was also expected since the higher the σ_{prior}, the less informative the prior is and therefore the higher the IG is from an acoustic observation.

3.4 Comparison of evaluated metrics

Comparison of the results is shown in Fig. 3. The diagonal line shows the histogram of the four metrics over the dataset. Correlating the metrics, it becomes clear that the largest correlation of > 0.99 is found between D and $\mathrm{IG}_{\mathrm{Baum}}$. Also, $\mathrm{IG}_{\mathrm{Baum}}$ and $\mathrm{IG}_{\mathrm{Barum}}$ correlate with a correlation coefficient of 0.64. SE does not correlate strongly with either of the other measures.

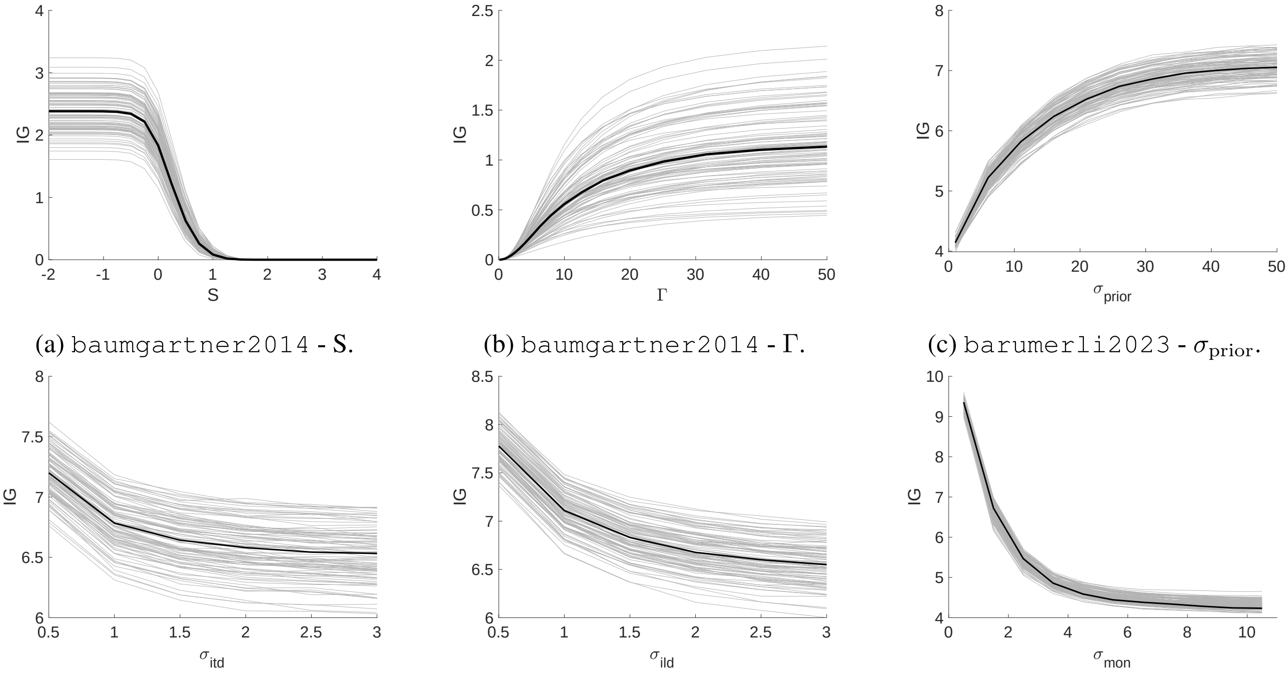
4. DISCUSSION

After computing the various measures, we now discuss the possible meaning of their magnitude and explain why discriminability and information gain for the `baumgartner2014` model correlate so strongly.

4.1 Magnitude of Information Content

Analyzing the magnitude of the outcomes, the most relevant result is the IG obtained for the `barumerli2014` model, which considers human perception and takes into account the complete sphere. There, the median information gain using default parameters is approximately 7.2 bits. One way of interpreting this result is that, assuming sound sources are equally likely from all directions, a human could *perfectly* distinguish $2^{7.2} \approx 147$ directions through listening. However, as we have shown, the absolute values strongly depend on the choice of parameters. Especially, reducing the monaural features noise σ_{mon} can easily increase the IG. Accordingly, if the absolute value is of interest, the model parameters first need to be fitted by means of perceptual data. Such fitted parameters would therefore provide meaningful insights about the information contained in the set of HRTFs.

The spectral entropy also takes the HRTFs on the entire sphere and does not have parameters. Here, the absolute number of bits was considerably lower than for $\mathrm{IG}_{\mathrm{Barum}}$. The spectral entropy only considers how much information is needed after compressing the HRTF matrix through linear operations. This



(a) baumgartner2014 - S. (b) baumgartner2014 - Γ. (c) barumerli2023 - σ_{prior}.

(d) barumerli2023 - σ_{itd}. (e) barumerli2023 - σ_{ild}. (f) barumerli2023 - σ_{mon}.

Figure 2: Influence of the model parameters S and Γ from baumgartner2014 and σ_{prior}, σ_{itd}, σ_{ild} and σ_{mon} from barumerli2023 on the IG. Each gray line represents an individual subject from the database. The black lines represents the median over subjects.

is far from what is thought to happen in human sound localization.

The IG results for baumgartner2014 with default parameters were below 1 bit for many parameter choices. Partly, this is explained by the model only considering positions in the median plane. However, intuitively, this result means that the listener would only be able to reliably tell apart less than two directions, which is lower than expected; when humans localize sound with their own HRTF, at least four quadrants are usually resolved. One possible cause might be parameterization. Choosing different parameters increased IG_{Baum} to up to 3 bits, which would allow listeners to distinguish eight directions in the median plane perfectly.

4.2 Discriminability vs. Information Gain

As described above, D and IG_{Baum} are highly correlated. This can easily be explained by the fact that they are both based on the same posterior distribution. For certain distributions, there is a direct relationship between variance and entropy. For example, if the posterior were a Gaussian distribution, there would be a direct relationship between variance and entropy, i.e., $H(X) = \frac{1}{2}\log_2(2\pi e\sigma^2)$. Thus, if the variance is low, the conditional entropy is also low, resulting in the IG in Eq. (5) to be high. While the exact law describing the relationship depends on the distribution, the high correlation between D and IG_{Baum} is therefore expected. Altogether, this shows that the ad-hoc metric introduced in [2] finds nearly the same differences as the information gain measure derived from more principled assumptions.

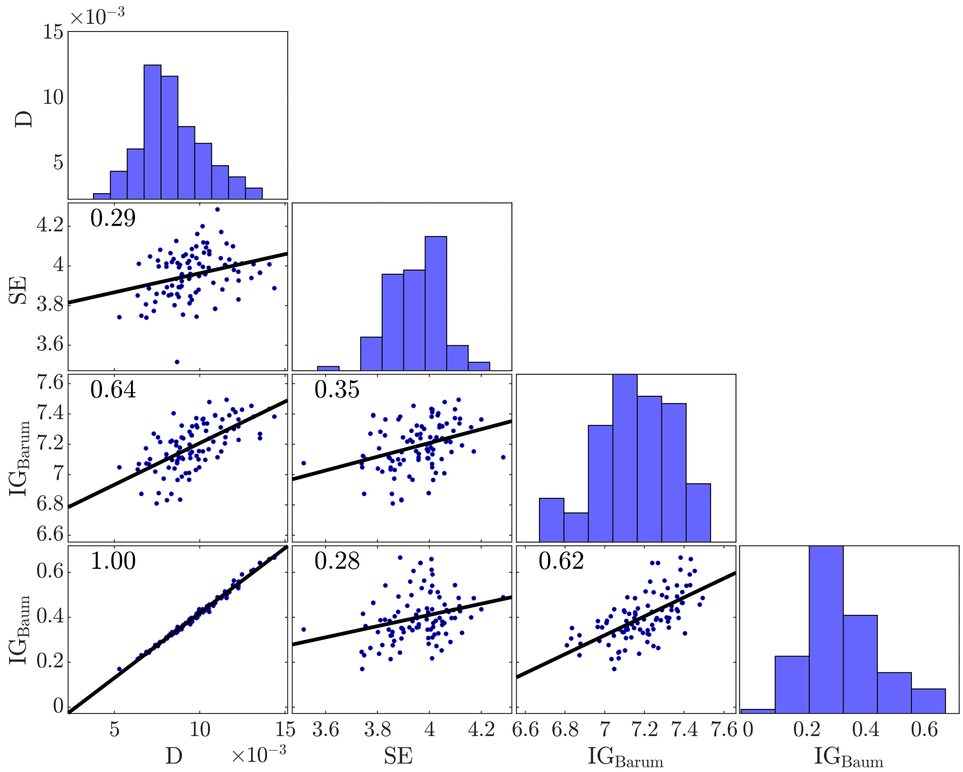


Figure 3: Correlation plot of the results. The diagonal shows histograms of the four measures, the scatter plots show correlations between them.

5. CONCLUSION

This paper compared three measures for quantifying the information contained in sets of HRTFs. We showed one model-independent measure, the spectral entropy, and two measures that are based on posterior distributions obtained from auditory models: the ad-hoc discriminability measure proposed in [2] and the information gain. We demonstrated that the magnitude of the results obtained from the model-based measures strongly depends on the model parameters, but that comparisons between sets of HRTFs are largely unaffected by the parameters.

We showed that when using the same model, discriminability and information gain are directly related. The spectral entropy did not correlate with either of the other two metrics. Since the information gain is well grounded in information theory and a Bayesian theory of localization, we propose to use it for selecting sets of HRTFs for a future adaptation study, in which it shall be assessed whether participants better adapt to informative non-individual HRTFs. However, as we have also pointed out, the information gain is contingent upon a specific auditory model, which must be carefully selected.

6. ACKNOWLEDGMENTS

This work was supported by EPSRC under grant EP/X032914/1.

7. REFERENCES

[1] C. Mendonça, "A review on auditory space adaptations to altered head-related cues," *Frontiers in Neuroscience*, vol. 8, Jul. 2014.

[2] P. Lladó, K. Pollack, and N. Meyer-Kahlen, "Towards a standard listener-independent HRTF to facilitate long-term adaptation," *J. Audio Eng. Soc*, vol. 72, pp. 188 – 192, Apr. 2024.

[3] V. Tourbabin and B. Rafaely, "The relation between the information delivered by head- related transfer function and human spatial hearing," in *Proceedings of Meetings on Acoustics*, May 2013.

[4] H. Peremans, G. McLachlan, P. Majdak, and J. Reijniers, "Ideal versus non-ideal observer models for sound localization," in *International Congress on Acoustics*, (Gyeongju, Korea), Oct. 2022.

[5] J. Reijniers, D. Vanderelst, C. Jin, S. Carlile, and H. Peremans, "An ideal-observer model of human sound localization," *Biological cybernetics*, vol. 108, pp. 169–181, 2014.

[6] J. Reijniers, G. McLachlan, B. Partoens, and H. Peremans, "Ideal-observer model of human sound localization of sources with unknown spectrum," *Scientific Reports*, vol. 15, no. 1, 2025.

[7] R. Barumerli, P. Majdak, M. Geronazzo, D. Meijer, F. Avanzini, and R. Baumgartner, "A bayesian model for human directional localization of broadband static sound sources," *Acta Acustica*, vol. 7, 2023.

[8] G. McLachlan, P. Majdak, J. Reijniers, and H. Peremans, "Towards modelling active sound localisation based on bayesian inference in a static environment," *Acta Acustica*, vol. 5, 2021.

[9] P. Lladó, R. Barumerli, R. Baumgartner, and P. Majdak, "Predicting the effect of headphones on the time to localize an auditory target," *Frontiers in Virtual Reality*, vol. 5, 2024.

[10] P. Majdak, C. Hollomey, and R. Baumgartner, "AMT 1. x: A toolbox for reproducible research in auditory modeling," *Acta Acustica*, vol. 6, 2022.

[11] R. Baumgartner, P. Majdak, and B. Laback, "Modeling sound-source localization in sagittal planes for human listeners," *J. Acoust. Soc. Am.*, vol. 136, pp. 791–802, Aug. 2014.

[12] R. Baumgartner, P. Majdak, and B. Laback, "Modeling Sound-Source Localization in Sagittal Planes for Human Listeners," *J. Acoust. Soc. Am.*, vol. 136, pp. 791–802, Aug. 2014.

[13] F. Brinkmann, M. Dinakaran, R. Pelzer, J. Wohlgemuth, F. Seipl, and S. Weinzierl, "The HUTUBS HRTF database," 2019. dx.doi.org/10.14279/depositonce-8487.

[14] O. Roy and M. Vetterli, "The Effective Rank: a Measure of Effective Dimensionality," in *5th European Signal Processing Conference*, Sept. 2007.

[15] R. Barumerli, P. Majdak, M. Geronazzo, D. Meijer, F. Avanzini, and R. Baumgartner, "A Bayesian model for human directional localization of broadband static sound sources," *Acta Acust.*, vol. 7, 2023.

[16] P. Lladó, P. Majdak, R. Barumerli, and R. Baumgartner, "Spectral weighting of monaural cues for auditory localization in sagittal planes," *Trends in Hearing*, vol. 29, 2025.