PROCEEDINGS of the
24th International Congress on Acoustics

October 24 to 28, 2022 in Gyeongju, Korea

ABS-0092

Ideal versus non-ideal observer models for sound localization

Herbert PEREMANS[(1)], Glen McLACHLAN[(2)], Piotr MAJDAK[(3)], Jonas REIJNIERS[(3)]

[(1)]Universiteit Antwerpen, Belgium, herbert.peremans@uantwerpen.be
[(2)]Universiteit Antwerpen, Belgium, glen.mclachlan@uantwerpen.be
[(3)]Acoustics Research Institute, Austria, piotr.majdak@oeaw.ac.at
[(3)]Universiteit Antwerpen, Belgium, jonas.reijniers@uantwerpen.be

ABSTRACT

Previously we derived an information theoretic measure to quantify sound localization performance complementing measures such as angular error and number of front-back confusions. It is based on Bayesian inference to derive the probability sound originates from particular directions given received acoustic input and prior information about possible source locations. In this model, we considered the joint probability of lateral and polar angles of the sound direction using ITD and both left and right monaural spectra as acoustic input. Ideally, all acoustic information is being considered when calculating this probability, in reality, a limited set of relevant features are extracted by the peripheral hearing system as determined by what is physically possible and what is sufficient for survival. Starting from this ideal-observer model as a baseline, we assess in this work to what extent the localization performance is affected by different heuristics. We show that the information-theoretic measure is well suited to study the impact of different sub-optimal processing strategies relying on increasingly simplified/reduced sets of features on localization performance. We propose that such an analysis can shed light on the trade-offs that give rise to non-ideal localization behavior by human listeners.

Keywords: Sound, Localization, Information loss

1 INTRODUCTION

In the ideal-observer model we derived previously [7], we determine the conditional probability distribution of the sound source direction $\boldsymbol{\theta}$, i.e., what is the probability that the sound originates from direction $\boldsymbol{\theta}$ given that we receive the acoustic input $\mathbf{A}$ and make use of a prior model M to interpret that acoustic input. Note that, assuming the sound source is in the far field, we only derive the direction of the sound source and ignore its distance from the listener.

Ideally, all acoustic information is being considered when inferring the sound direction. This entails registering and processing both sound waves arriving at the ears. In reality, though, we know that auditory system processes can 'measure' these signals with some restrictions. A number of 'relevant' features are extracted from the acoustic signals and transmitted for further processing, where feature space is determined by what is physically possible and, given our evolutionary history, what is necessary and sufficient for survival.

In our original model formulation, we assumed that this feature space consisted of the interaural time difference (ITD) and (the log-magnitude of) the spectrum of both the left and right monaural acoustic input, i.e. Head Related Transfer Functions (HRTFs), as measured by a filterbank of bandpass filters. This choice was made because these features are either 'measured' directly by the cochlea or can be derived directly from the output of the cochlea in the case of ITD. Moreover, the resolution with which these features are measured (in terms of time resolution for ITD, spectral resolution with respect to the number of independent frequency channels and the resolution of magnitude in each of the channels) have been experimentally determined.

The optimal way of processing the information contained in these features, which are transmitted with uncertainty due to the limitations of the auditory system, is through Bayesian inference. In this sense, the model we proposed in [7] is an ideal-observer model. It does not make any assumptions on how information is actually processed by the brain, it should be understood rather as an upper limit on what is theoretically feasible given the acoustic information and the limitations of the auditory system. Nevertheless, the qualitative predictions from this ideal-observer model are in good agreement with actual human sound localization.

In this work, we study how limiting the observer model to various subsets of the above mentioned features, i.e. replacing the ideal-observer model by various non-ideal ones, affects localization performance. To quantify performance we report both classic performance measures such as angular errors as well as a less widely used mutual information measure [8]. The advantage of angular errors is that this performance measure can be compared directly to human localization experiments. The main advantage of the mutual information performance measure is that, contrary to angular errors, it is agnostic with respect to the decision rule used to turn the posterior distribution of the source direction into an actual estimate. In this study we use a 'maximum a posteriori probability' (MAP) decision rule. As the actual decision rule used by human listeners is still the subject of an ongoing debate [2], the mutual information measure should prove a more stable measure of performance.

2 MODELLING HUMAN SOUND LOCALIZATION

2.1 Ideal-observer model

As mentioned above, we base our results on the calculation of the conditional posterior probability distribution of the source direction $\boldsymbol{\theta}$

$$p(\boldsymbol{\theta}|\mathbf{A},M), \tag{1}$$

with the internal model M (see [7] for details) conflating all prior knowledge defined by

1. the mappings describing the direction-dependent acoustic features (ITD($\boldsymbol{\theta}$) and log-magnitude of the binaural HRTF $\mathbf{H_L}(\boldsymbol{\theta})$ and $\mathbf{H_R}(\boldsymbol{\theta})$) as encoded in the memorised templates (see below),
2. measurement noise on the ITD and spectral logmagnitudes (σ_{itd} and $\boldsymbol{\Sigma}_I$),
3. uncertainty of (the log-magnitude of) the source spectrum $p(\boldsymbol{S}) \sim N(\boldsymbol{S}_0, \boldsymbol{\Sigma}_S)$,
4. uncertainty of the initial source position $p_0(\boldsymbol{\theta})$ which, in the absence of any prior information about possible locations of the source would be a uniform distribution defined over the sphere centered on the listener's head.

Without loss of information we consider the following representation of the acoustic input:

$$\mathbf{A} = [itd, \mathbf{S_L} - \mathbf{S_R}, \frac{1}{2}(\mathbf{S_L} + \mathbf{S_R})] := [\mathbf{A_0}, \mathbf{A_S}], \tag{2}$$

with $\mathbf{S_L}$ and $\mathbf{S_R}$ the log-magnitude of the left and right ear spectra respectively. This makes, as the (unknown) source spectrum is in both the left and right ear spectral input, that the source spectrum is only present in $\mathbf{A_S} = \frac{1}{2}(\mathbf{S_L} + \mathbf{S_R})$. Consequently, the mappings describing the direction-dependent acoustic features can be encoded in the following template set,

$$\mathbf{T}(\boldsymbol{\theta}) = [itd_T(\boldsymbol{\theta}), \mathbf{H_L}(\boldsymbol{\theta}) - \mathbf{H_R}(\boldsymbol{\theta}), \frac{1}{2}(\mathbf{H_L}(\boldsymbol{\theta}) + \mathbf{H_R}(\boldsymbol{\theta})) + \boldsymbol{S}_0] := [\mathbf{T_0}(\boldsymbol{\theta}), \mathbf{T_S}(\boldsymbol{\theta})]. \tag{3}$$

In many psycho-acoustic experiments interaural time difference (ITD) and interaural level difference (ILD) cues are studied as lateralization cues. We will assume that the listener, when using ILD cues, simplifies the processing by reducing the actual pattern of cochlear activity to the summed cochlear activity. Hence, we define the ILD as $ild = \sum_{f=f_{min}}^{f_{max}} S_L(f) - S_R(f)$. Please note that other more perceptually accurate definitions could be used instead.

Different from [7] we define the source direction $\boldsymbol{\theta}$ not in terms of the joint distribution of azimuth and elevation angles. Instead, we determine first the distribution of the lateral angle $\theta_l \in [-\frac{\pi}{2}, \frac{\pi}{2}]$ of the source direction $p(\theta_l|\mathbf{A},M)$ and subsequently the conditional distribution of the polar angle $\theta_p \in [-\pi, \pi]$ given the lateral angle $p(\theta_p|\theta_l,\mathbf{A},M)$ using all available acoustic cues $\mathbf{A}$. We can then derive the joint distribution by

$$p(\boldsymbol{\theta}|\mathbf{A},M) = p(\theta_l|\mathbf{A},M) \cdot p(\theta_p|\theta_l,\mathbf{A},M), \tag{4}$$

We propose that it is useful to split the joint distribution in this way as we can now more readily study the effects of non-ideal observer models using different subsets of all available acoustic features for determining lateral and polar angles.

2.2 Mutual Information

We consider a localization experiment wherein during trial i a source is presented at target direction $z = \boldsymbol{\theta}^T$. We define the information I extracted by a listener from the acoustic stimuli $\boldsymbol{a}_i$ received during this localization experiment by the mutual information between the acoustic stimuli and the source direction [1]. This measure quantifies the average reduction of uncertainty about the source direction that occurs in the course of one trial of such a localization experiment

$$I(\boldsymbol{\theta};\boldsymbol{a}=\boldsymbol{a}_i|z=\boldsymbol{\theta}^T) = H(\boldsymbol{\theta}|z=\boldsymbol{\theta}^T) - H(\boldsymbol{\theta}|\boldsymbol{a}=\boldsymbol{a}_i, z=\boldsymbol{\theta}^T), \tag{5}$$

with

$$H(\boldsymbol{\theta}|z=\boldsymbol{\theta}^T) = -\sum_{k=1}^{n} p(\boldsymbol{\theta}_k|z=\boldsymbol{\theta}^T)\log_2 p(\boldsymbol{\theta}_k|z=\boldsymbol{\theta}^T) \tag{6}$$

and

$$H(\boldsymbol{\theta}|\boldsymbol{a}=\boldsymbol{a}_i, z=\boldsymbol{\theta}^T) = -\sum_{k=1}^{n} p(\boldsymbol{\theta}_k|\boldsymbol{a}=\boldsymbol{a}_i, z=\boldsymbol{\theta}^T)\log_2 p(\boldsymbol{\theta}_k|\boldsymbol{a}=\boldsymbol{a}_i, z=\boldsymbol{\theta}^T) \tag{7}$$

being the entropy of the prior and posterior distribution of the source direction respectively expressed in bits. Note that we assume a discrete sampling grid $\boldsymbol{\theta}_k$ with k ranging from 1 to n. A uniform distribution would therefore have an entropy of $H = \log_2(n)$ bits.

By approximating the expected value over the acoustic input $\boldsymbol{a}$ as generated by a source at target direction $z = \boldsymbol{\theta}^T$ with a Monte Carlo estimate over m trials

$$H(\boldsymbol{\theta}|\mathbf{A}, z=\boldsymbol{\theta}^T) \cong \sum_{i=1}^{m} p(\boldsymbol{a}_i|z=\boldsymbol{\theta}^T)\cdot H(\boldsymbol{\theta}|\boldsymbol{a}=\boldsymbol{a}_i, z=\boldsymbol{\theta}^T),$$

we can write the mutual information between $\boldsymbol{\theta}$ and $\mathbf{A}$ given the target direction $z = \boldsymbol{\theta}^T$ as

$$I(\boldsymbol{\theta};\mathbf{A}|z=\boldsymbol{\theta}^T) = H(\boldsymbol{\theta}|z=\boldsymbol{\theta}^T) - H(\boldsymbol{\theta}|\mathbf{A}, z=\boldsymbol{\theta}^T). \tag{8}$$

It measures the expected reduction in uncertainty/entropy about the source direction that results from learning the value of the acoustic input for a target direction $z = \boldsymbol{\theta}^T$.

By replacing the joint distribution from Eq. 1 by the split version of Eq. 4, the mutual information (Eq.8) can also be split in the information contained in the acoustic cues about the lateral angle and the information contained in the acoustic cues about the polar angle given the lateral angle

$$\begin{aligned}
I(\boldsymbol{\theta};\mathbf{A}|z=\boldsymbol{\theta}^T) &= H(\boldsymbol{\theta}|z=\boldsymbol{\theta}^T) - H(\boldsymbol{\theta}|\mathbf{A}, z=\boldsymbol{\theta}^T)\\
&= H(\theta_l|z=\boldsymbol{\theta}^T) - H(\theta_l|\mathbf{A}, z=\boldsymbol{\theta}^T) + H(\theta_p|\theta_l, z=\boldsymbol{\theta}^T) - H(\theta_p|\theta_l, \mathbf{A}, z=\boldsymbol{\theta}^T)\\
&= I(\theta_l;\mathbf{A}|z=\boldsymbol{\theta}^T) + I(\theta_p;\mathbf{A}|\theta_l, z=\boldsymbol{\theta}^T).
\end{aligned}$$

2.3 Non-ideal observer models

Our brain may not use the optimal Bayesian strategy, because there may be other (less ideal with regards to accuracy) ways of processing the acoustic information that are more robust or allow for a more efficient calculation, and are ‘good enough’ for the task at hand. Also, some of the cues might be more or less informative depending on the situation, e.g. knowledge of the source spectrum. Various authors [4, 6, 3] have studied to what extent lateral and polar angles of the source direction are determined based on certain subsets of acoustic cues only. From a Bayesian inference point of view all such schemes are suboptimal. Indeed, as not all information contained in the cues is used this results in approximations of varying accuracy

$$p(\boldsymbol{\theta}|\mathbf{A}, M) \approx p(\theta_l|\mathbf{A_l}, M)\cdot p(\theta_p|\theta_l, \mathbf{A_p}, M), \tag{9}$$

depending on the particular choice of the subsets of acoustic cues $\mathbf{A_l}$ and $\mathbf{A_p}$ used to determine the lateral and polar angles respectively.

In the following, we investigate the impact of various suboptimal processing strategies, i.e. heuristics, that rely on particular subsets of features. We use the ideal-observer model as a baseline, and then assess to what extent the localization performance deteriorates for these heuristics. The associated information loss is defined by

$$\begin{aligned} I_{loss} &= I(\boldsymbol{\theta};\mathbf{A}|z=\boldsymbol{\theta}^T) - (I(\theta_l;\mathbf{A_l}|z=\boldsymbol{\theta}^T) + I(\theta_p;\mathbf{A_p}|\theta_l,z=\boldsymbol{\theta}^T)) \\ &= I(\theta_l;\mathbf{A}|z=\boldsymbol{\theta}^T) - I(\theta_l;\mathbf{A_l}|z=\boldsymbol{\theta}^T) + I(\theta_p;\mathbf{A}|\theta_l,z=\boldsymbol{\theta}^T) - I(\theta_p;\mathbf{A_p}|\theta_l,z=\boldsymbol{\theta}^T) \\ &= I^{lateral}_{loss} + I^{polar}_{loss} \geq 0. \end{aligned} \tag{10}$$

3 RESULTS

The HRTFs used for the calculations in this section were from the ARI(ALTB)-database. Unless indicated otherwise, the results were averaged over 10 subjects from the database that were randomly selected.

The simulated experiment consists of $N = 100$ trials for each potential source position. The results are the average over these N trials. Lateral and polar angle errors are defined as the absolute value of the difference between the true source lateral and polar angle and their maximum a posteriori probability (MAP) estimates as derived from the posterior probability distribution $p(\boldsymbol{\theta}|\mathbf{A},M)$. Front-back confusions are registered whenever the angle between the estimated source direction and the mirror-image (re. to frontal plane) of the sound source is smaller than for the true source direction.

In our original ideal-observer model we had the listener localize the source by jointly estimating azimuth and elevation angles. However, it is more natural to express the source direction in terms of lateral and polar angles that need not necessarily be estimated based on the same cues. Often ITD and ILD cues are considered most important for lateralization and spectral cues for determining the polar angle. Below we investigate to what extent such a division can be explained by a rational use of the information contained in the cues.

3.1 Feature spaces for lateral angle estimation

To determine the contributions to lateralization of the different acoustic cues, we study the loss of information associated with determining the lateral angle based on increasingly smaller subsets of the available acoustic cues. Hence, we determine how the uncertainty as encoded by the entropy of the probability distribution $p(\theta_l|\mathbf{A_l},M)$ (see Eq. 9) increases with smaller subsets of acoustic features $\mathbf{A_l}$. The results are shown in Fig. 1. Note that the regions of information loss correspond to the regions where lateral angular error is higher as does the amount of information loss with the degree of accuracy loss. Note however, that both the size and the spatial distribution of the errors will change with the decision rule (MAP) whereas the mutual information loss only depends on the HRTFs of the listeners.

As the source spectrum $\mathbf{S}_0$ or its variability is not always known, it can be advantageous for the listener to limit oneself to acoustic cues $\mathbf{A}_0$ (Eq. 2) that are independent of the source spectrum. In the top two rows of Fig. 1, we show how using ITD and spectral difference cues compares with the use of all acoustic cues for the determination of the lateral angle. This comparison shows clearly that very little angular information and hence accuracy is lost. Fig. 2(a) shows the information loss I_{loss} due to using only source spectrum independent cues instead of all cues. Lateral angle information is lost mostly along the midsaggital plane. Note that this information loss increases Ls the uncertainty on the source spectrum decreases because spectral cues contain more information in that case. Hence, we conclude that robustness with respect to an unknown source spectrum can be gained at the price of a small information loss along the midsaggital plane by deriving the lateral angle estimate from $[itd,\mathbf{S_L}-\mathbf{S_R}]$-cues only. Fig. 2(b-d) also illustrates that there is quite a bit of variability underlying these results depending on the actual listeners' HRTFs possibly explaining the large variability in experimental subjects' performance.

Rows 1 and 3 in Fig. 1 show how combining ITD and ILD cues compares with the use of all acoustic cues for the determination of the lateral angle of the sound source. We observe that in this case while larger angular errors occur, lateralization is still quite good indicating that for lateral angle determination detailed spectral

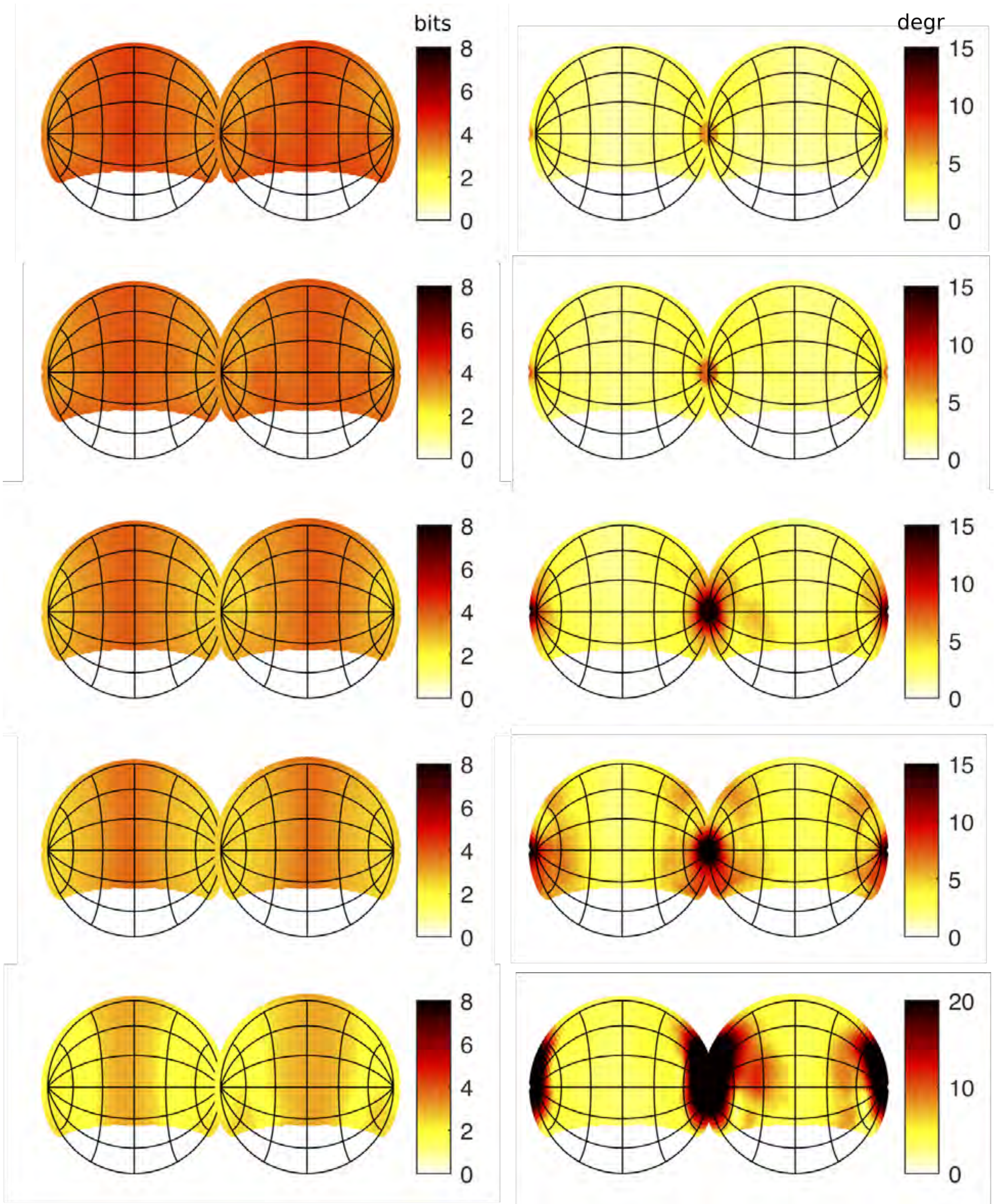

Figure 1. Determining the lateral angle of the sound source direction. Left column: mutual information; Right column: lateral angle error. From top to bottom: all cues used; $[itd, \mathbf{S_L} - \mathbf{S_R}]$-cues used; $[itd, ild]$-cues used;$[itd]$-cues used;$[ild]$-cues used.

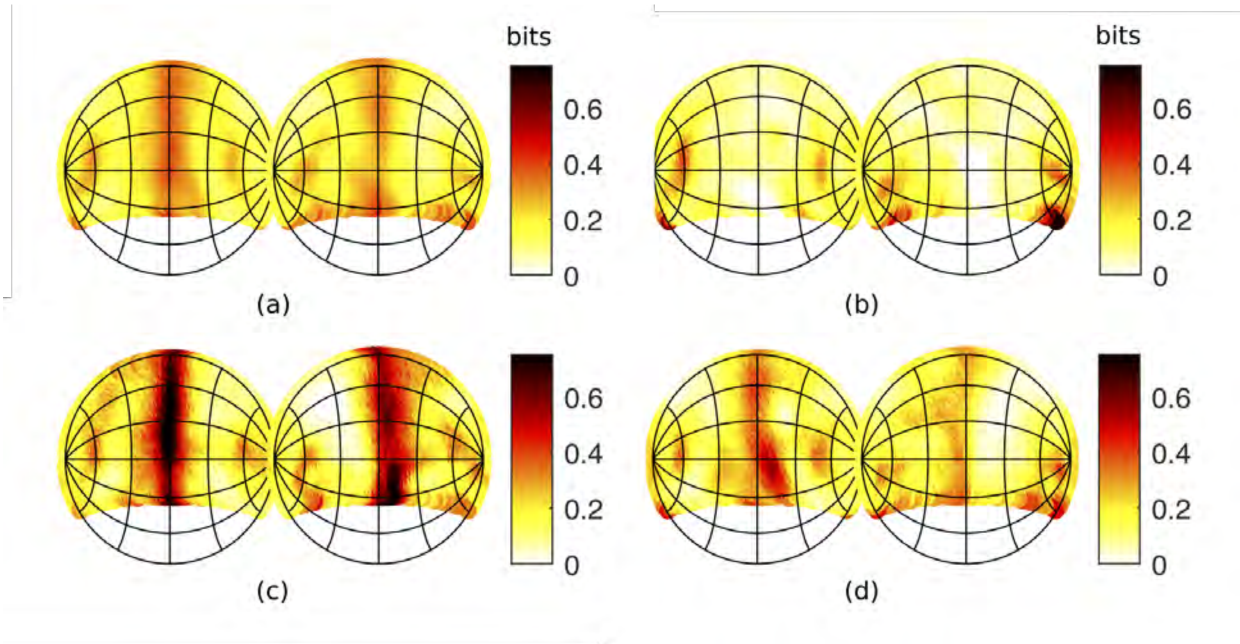

Figure 2. Information loss due to using only $[itd, \mathbf{S_L} - \mathbf{S_R}]$-cues; (a) all listeners; (b) listener 7; (c) listener 4; (d) listener 1.

information helps but is not essential. From Fig. 1, most information loss I_{loss} occurs on the sides where we also observe the largest angular accuracy decrease. Hence, apart from a small contribution along the midsaggital plane (see above), spectral cues make the largest contribution to the lateral angle estimate in these peripheral regions. This effect can be observed even more clearly when estimating the lateral angle on either the ILD of the ITD cue alone.

In accordance to the duplex theory of sound localization listeners rely on ITD cues for low-pass stimuli and ILD cues for high-pass stimuli and, in addition, place a higher weight on ITD cues for wideband stimuli [4]. In the lower two rows of Fig. 1 we include the results for using ITD cues only and ILD cues only when estimating the lateral angle of the sound source. When comparing the $[itd, ild]$-cues strategy with the $[itd]$-cue only strategy (rows 3 and 4) we note that by exploiting ILD-cues on top of ITD-cues information is mostly gained for the $\theta_l > 30^\circ$ region. A similar conclusion can be drawn when comparing the $[itd, ild]$-cues strategy with the $[ild]$-cue only strategy (rows 3 and 5). But the information loss is clearly more pronounced when relying on ILD cues only compared to ITD cues only. This shows that the higher weight placed by listeners on ITD cues for wideband stimuli, i.e. when both cues are present, can indeed be a rational strategy.

3.2 Feature spaces for polar angle estimation

To verify the claim that spectral cues are the dominant cues for polar angle estimation, we study the loss of information associated with determining the polar angle, given the lateral angle, based on different subsets of the available acoustic cues. We determine how the entropy of the probability distribution $p(\theta_p|\theta_l, \mathbf{A_p}, M)$ (see Eq. 9) depends on the subsets of acoustic features $\mathbf{A_p}$. From a comparison of the results shown in Fig. 3 we conclude that, as expected, no information is lost by determining the polar angle based on spectral cues only, given the lateral angle. In Fig. 3 we also show the effect of using selected monaural cues, i.e. the listener relies on right ear spectral cues in the right hemisphere and left ear spectral cues in the left hemisphere only to determine the polar angle of the sound source. Such a strategy would clearly result in a significant increase in information loss as well as concomitant polar angle errors for all directions. We can compare this strategy with the binaural weighting scheme described in [5], where it was found that the monaural estimates of the polar

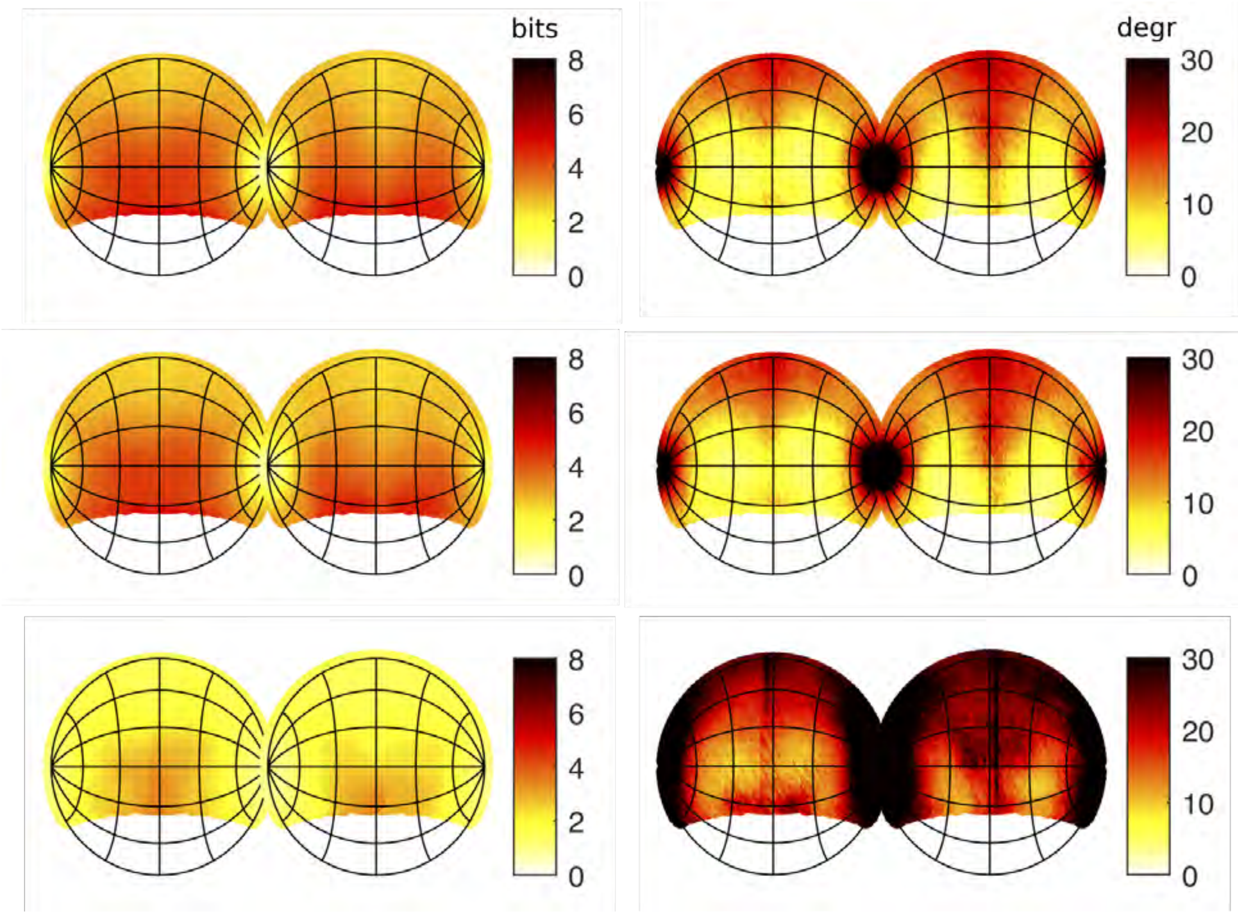

Figure 3. Determining the polar angle (given the lateral angle) of the sound source direction. Left column: mutual information; Right column: polar angle error. From top to bottom: $[itd, \mathbf{S_L} - \mathbf{S_R}, \frac{1}{2}(\mathbf{S_L} + \mathbf{S_R})]$-cues used; $[\mathbf{S_L} - \mathbf{S_R}, \frac{1}{2}(\mathbf{S_L} + \mathbf{S_R})]$-cues used; $[(\theta_l >= 0) \cdot \mathbf{S_L} + (\theta_l < 0) \cdot \mathbf{S_R}]$-cues used.

angle could be understood as a weighted sum of estimates based on the ipsilateral ear spectrum (weight=1.0) and the contralateral ear spectrum (weight <0.2) for source positions in the frontal hemisphere ($|\theta_l| >= 30°$) and equal weights in the midsaggital plane. It seems that ours is an extreme version of such a weighting scheme and further study is required to understand how making it more realistic would affect the results shown here.

4 CONCLUSIONS

In this study we investigated the use of mutual information as a powerful way to analyse performance of various suboptimal strategies of processing the acoustic cues available to a listener when estimating the direction of a sound source. We propose that by analysing the spatial distributions of the information loss occurring for the different strategies further insight can be gained in the trade-offs, e.g. robustness or speed vs. accuracy, faced by the listener's brain. In particular, the results of our analysis provide an information-theoretic argument in support of the hypothesis that rational listeners derive lateral angle information from binaural, source spectrum independent, cues and polar angle from binaural spectral cues. In addition, we propose that, as regions of high information loss correlate well with areas of higher angular errors and as it is independent of a specific decision rule, mutual information provides a stable measure of performance.

ACKNOWLEDGEMENTS

This research was supported by the Research Foundation Flanders (FWO) under Grant nr. G023619N and by the Agency for Innovation and Entrepreneurship (VLAIO). P.M. received funding from the European Union's Horizon 2020 research and innovation funding programme (grant no. 101017743, project "SONICOM"–Transforming auditory-based social interaction and communication in AR/VR).

REFERENCES

[1] T. M. Cover and J. A. Thomas. *Elements of Information Theory*. Wiley, 2nd edition, 2006.

[2] R. Ege, A. J. Van Opstal, and M. M. Van Wanrooij. Accuracy-precision trade-off in human sound localisaion. *Sci Rep 8*, page 16399, 2018.

[3] P. M. Hofman and J. A. Van Opstal. Spectro-temporal factors in two-dimensional human sound localization. *The Journal of the Acoustical Society of America*, 103(5):2634–2648, 1998.

[4] E. A. Macpherson and J. C. Middlebrooks. Listener weighting of cues for lateral angle: The duplex theory of sound localization revisited. *The Journal of the Acoustical Society of America*, 111(5):2219–2236, 2002.

[5] E. A. Macpherson and A. T. Sabin. Binaural weighting of monaural spectral cues for sound localization. *The Journal of the Acoustical Society of America*, 121(6):3677–3688, 2007.

[6] M. Morimoto and H. Aokata. Localization cues of sound sources in the upper hemisphere. *Journal of the Acoustical Society of Japan (E)*, 5(3):165–173, 1984.

[7] J. Reijniers, D. Vanderelst, C. Jin, S. Carlile, and H. Peremans. An ideal-observer model of human sound localization. *Biological cybernetics*, 108(2):169–181, 2014.

[8] J. Reijniers, D. Vanderelst, and H. Peremans. Morphology-induced information transfer in bat sonar. *Phys. Rev. Lett.*, 105:148701, Sep 2010.