Acta Acustica 2023, **7**, 12
© The Author(s), Published by EDP Sciences, 2023
https://doi.org/10.1051/aacus/2023006

Available online at:
https://acta-acustica.edpsciences.org

Topical Issue - Auditory models: from binaural processing to multimodal cognition

Scientific Article   Open Access

A Bayesian model for human directional localization of broadband static sound sources

Roberto Barumerli[1,*], Piotr Majdak[1], Michele Geronazzo[2,3], David Meijer[1], Federico Avanzini[4], and Robert Baumgartner[1]

[1] Acoustics Research Institute, Austrian Academy of Sciences, 1040 Vienna, Austria
[2] Department of Management and Engineering, University of Padova, 36100 Vicenza, Italy
[3] Dyson School of Design Engineering, Imperial College London, London, United Kingdom
[4] Department of Computer Science, University of Milano, 20133 Milan, Italy

Received 5 June 2021, Accepted 13 March 2023

Abstract – Humans estimate sound-source directions by combining prior beliefs with sensory evidence. Prior beliefs represent statistical knowledge about the environment, and the sensory evidence consists of auditory features such as interaural disparities and monaural spectral shapes. Models of directional sound localization often impose constraints on the contribution of these features to either the horizontal or vertical dimension. Instead, we propose a Bayesian model that flexibly incorporates each feature according to its spatial precision and integrates prior beliefs in the inference process. The model estimates the direction of a single, broadband, stationary sound source presented to a static human listener in an anechoic environment. We simplified interaural features to be broadband and compared two model variants, each considering a different type of monaural spectral features: magnitude profiles and gradient profiles. Both model variants were fitted to the baseline performance of five listeners and evaluated on the effects of localizing with non-individual head-related transfer functions (HRTFs) and sounds with rippled spectrum. We found that the variant equipped with spectral gradient profiles outperformed other localization models. The proposed model appears particularly useful for the evaluation of HRTFs and may serve as a basis for future extensions towards modeling dynamic listening conditions.

Keywords: Sound localization, Auditory model, Bayesian inference, HRTF, Behavior

1 Introduction

When localizing a sound source, human listeners have to deal with numerous sources of uncertainty [1]. Uncertainties originate from ambiguities in the acoustic signal encoding the source position [2] as well as the limited precision of the auditory system in decoding the received acoustic information [3, 4]. Bayesian inference describes a statistically optimal solution to deal with such uncertainties [5] and has been applied to model sound localization in various ways [6–9].

Typical approaches of sound localization models rely on the evaluation of several spatial auditory features. Head-related transfer functions (HRTFs) describe the spatially dependent acoustic filtering produced by the listener's ears, head, and body [10] and have been used to derive spatial auditory features. The way to extract those features is a matter of debate. In particular, a large variety of monaural spectral-shape features have been studied [11–17], with spectral magnitude profiles [14, 17] and spectral gradient profiles [12, 15] being the most established ones. Despite such details, there is consensus that the interaural time and level differences (ITDs and ILDs) [1] as well as some form of monaural spectral shapes are important features for the directional localization of broadband sound sources [18].

In order to decode the spatial direction from the auditory features, models rely on the assumption that listeners have learned to associate acoustic features with spatial directions [13, 19]. Interaural features are particularly informative about the lateral directions (left/right) and ambiguous with respect to perpendicular directions along sagittal planes (top/down and front/back) for which the monaural spectral features are more informative [18]. This phenomenon led to propose a variant of the spherical coordinate system, the so-called interaural-polar coordinate system where the two poles are placed on the interaural axis [20]. In this article, we use the so-called modified interaural-polar coordinate system, with the lateral angle describing a source along the lateral directions ranging from −90° (left)

[*] Corresponding author: roberto.barumerli@oeaw.ac.at

to 90° (right) and the polar angle describing a source along a sagittal plane in the interval ranging from $-90°$ (bottom), via 0° (eye level, front), 90° (top), and 180° (back) to 270° [21]. Directional sound-localization studies typically use the interaural-polar coordinate system to separate the effects of the interaural and monaural features [22, 23]. However, this separation is a simplification. For example, monaural spectral features can also contribute to the direction estimation along the lateral dimension [24–26]. Hence, directional sound-localization models may provide better predictions when jointly exploiting the information encoded by all auditory features.

Such joint information has already been considered in a model of directional sound localization based on Bayesian inference [6]. This model computes a spatial likelihood function from a precision-weighted integration of noisy acoustic features. Then, the perceived source direction is assumed to be at the maximum of that likelihood function. While this model was built to assess which spatial information can be accessible to the auditory system, its predictions overestimate the actual human performance yielding unrealistically low front-back confusion rates and localization errors [27]. Still, to model human performance, this model can serve as a solid basis for improvements such as the consideration of monaural spectral features, the integration of response noise involved in typical localization tasks, and the incorporation of prior beliefs.

Prior beliefs are essential in the process of Bayesian inference because they reflect the listener's statistical knowledge about the environment, helping to compensate for uncertainties in the sensory evidence [28]. For example, listeners seem to effectively increase precision in a frontal localization task by assuming source directions to be more likely located at the eye-level rather than at extreme vertical positions [8]. However, such an increase in precision may come at the cost of decreasing accuracy. As it seems, the optimal accuracy-precision trade-off in directional localization depends on the statistical distribution of sound sources [29]. While listeners seem to adjust their prior beliefs to changes in the sound-source distribution [29, 30], they may also establish long-term priors reflecting the distribution of sound sources in their everyday environment.

Here, we introduce a Bayesian inference model to predict the performance of a listener estimating the direction of static broadband sounds. The model implements a noisy feature extraction and probabilistically combines interaural and monaural spatial features [6]. We limit our model to an anechoic auditory scene with a single broadband and stationary sound source, without listener and source movements. In this scenario, the representation of monaural features requires to account for spectral information [8, 15] while interaural features computation can rely on broadband estimators [18, 31]. Despite such simplifications, the model structure can easily include more complex processing of interaural features as required for narrow-band stimuli or reverberant and multisource environments (e.g. [32]). Our model computes the likelihood function by comparing the features with templates (i.e. spatial features obtained from listener-specific HRTFs). Subsequently, the probabilistic representation of the sound direction results from combining the sensory evidence with prior beliefs. For simplicity, we consider static prior emphasizing directions at the eye level [8], while the model structure can easily integrate future extensions towards more flexible, task-dependent priors. A Bayesian decision function estimates the source position from the resulting spatial representation. As a last step, the model incorporates response scattering to account for the uncertainty introduced by pointing response in localization experiments [15].

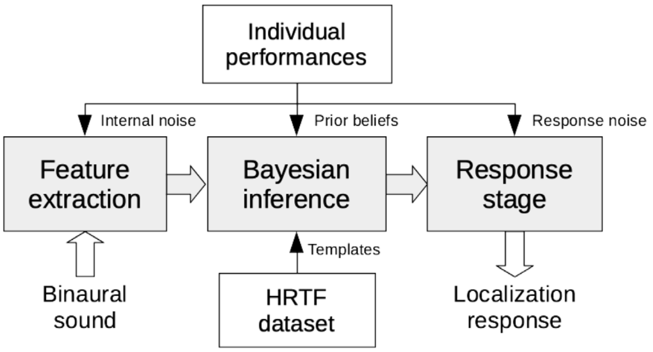

Figure 1. Model structure. Gray blocks: Model's processing pipeline consisting of 1) the *feature extraction* stage to compute spatial features from the binaural sound; 2) the *Bayesian inference* stage integrating the sensory evidence obtained by comparison with feature templates and prior beliefs to estimate the most probable direction; and 3) the *response stage* transforming the internal estimate to the final localization response. White blocks: Elements required to fit the model to an individual subject consisting of listener performances in estimating sound direction and individual HRTF dataset.

For evaluation, we considered a model variant based on spectral amplitudes and a model variant based on spectral gradients [6]. The model's parameters were fitted to the sound-localization performance of individual listeners [23]. We then tested the simulated responses of both model variants against human responses from sound-localization experiments investigating the effects of non-individual HRTFs [22] and ripples in the source spectrum [33].

The paper is organized as follows: Section 2 describes the auditory model (Sect. 2.1) and explains the parameter estimation (Sect. 2.2). Then, Section 3 evaluates the model's performance by comparing its estimations to the actual performance of human listeners. Finally, Section 4 discusses the model's relevance as well as its limitations, and outlines its potential for future extensions.

2 Methods

2.1 Model description

The proposed auditory model consists of three main stages, as shown in Figure 1: 1) The feature extraction stage determines the encoded acoustic spatial information represented as a set of spatial features altered by noise; 2) The Bayesian inference integrates the sensory evidence resulting from the decoding procedure based on feature templates with the prior belief and forms a perceptual decision; and 3) The response stage transforms the perceptual decision in a directional response by corrupting the estimation with uncertainty in the pointing action.

2.1.1 Feature extraction

The directional transfer function transformed in the time domain (i.e. the HRTF processed to remove the direction-independent component [34]) convolved with the sound source provides the sensory evidence, which is represented by the spatial auditory features. We follow [6] in that we decode the spatial information provided by a single sound source via the binaural stimulus from a vector of spatial features:

$$\bar{\boldsymbol{t}} = [x_{\text{itd}}, x_{\text{ild}}, \boldsymbol{x}_{\text{L,mon}}, \boldsymbol{x}_{\text{R,mon}}], \tag{1}$$

where x_{itd} denotes a scalar ITD feature, x_{ild} a scalar ILD feature, and a vector that concatenates monaural spectral features for left ear, $\boldsymbol{x}_{\text{L,mon}}$, and right ear, $\boldsymbol{x}_{\text{R,mon}}$. Each feature is assumed to be extracted by different neural pathways responsible to deliver encoded spatial information to higher levels of the auditory system [1, 4].

Simple broad-band estimators approximated the interaural features [18, 31] because we limited our evaluation to the task of localizing a broadband and spatially static sound source in an acoustic free field. More complex representations of interaural features are required when considering more natural listening conditions, e.g., reverberant spaces, multi-talker environments, sounds embedded in noise, or sounds with spectro-temporal fluctuations such as speech or music. In our simple listening scenario, the ILD was approximated as the time-averaged broadband level difference between the left and right channels [18]. The ITD was estimated by first processing each channel of the binaural signal with a low-pass Butterworth filter (10th order and cutoff 3000 Hz) and an envelope extraction step based on the Hilbert transform. Then, the ITD was computed with the interaural cross-correlation method which is a good estimator of perceived lateralization in static scenarios with noise bursts [31]. In addition, we applied the transformation proposed by Reijniers *et al.* [6] to compensate the increasing uncertainty levels for increasing ITDs [35] resulting in a dimensionless quantity with a more isotropic variance:

$$x_{\text{itd}} = \frac{\operatorname{sgn}(\text{itd})}{b_{\text{itd}}} \log\left(1 + \frac{b_{\text{itd}}}{a_{\text{itd}}} \cdot |\text{itd}|\right), \tag{2}$$

with "itd" denoting ITDs in µs and the parameters $a_{\text{itd}} = 32.5$ µs and $b_{\text{itd}} = 0.095$ and "sgn" indicating the sign function (for details on the derivation based on signal detection theory, see Supplementary Information from [6]). An example of the interaural features as functions of the lateral angle is shown in Figure 2.

Monaural spectral features, $\boldsymbol{x}_{\{\text{L,R}\},\text{mon}}$, were derived from approximate neural excitation patterns. To approximate the spectral resolution of the human cochlea, we processed the binaural signal by the gammatone filterbank with non-overlapping equivalent rectangular bandwidths [36, 37], resulting in $N_{\text{B}} = 27$ bands within the interval [0.7, 18] kHz [38, 39]. Followed by half-wave rectification and square-root compression to model hair-cell transduction (e.g., [32, 40]), it resulted in the unit-less excitation:

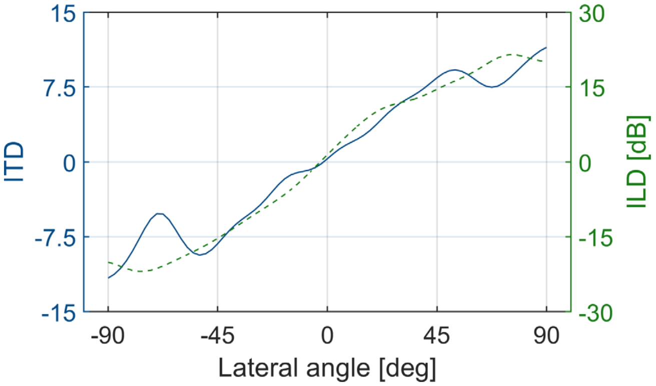

Figure 2. Interaural features as functions of lateral angle in the horizontal frontal plane. Left axis (blue solid line): Transformed ITD x_{itd} (dimensionless), see equation (2). Right axis (green dashed line): ILD (in dB) obtained from the magnitude profiles. Example for subject NH12 [23].

$$\bar{c}^{\boldsymbol{\varphi}}_{\zeta,b}[n] = (h^{\boldsymbol{\varphi}}_{\zeta} * g_b)[n],$$

$$c^{\boldsymbol{\varphi}}_{\zeta,b}[n] = \begin{cases} \sqrt{\bar{c}^{\boldsymbol{\varphi}}_{\zeta,b}[n]} & \text{if } \bar{c}^{\boldsymbol{\varphi}}_{\zeta,b}[n] \geq 0 \\ 0 & \text{otherwise,} \end{cases} \tag{3}$$

where subscripts $\zeta \in \{L, R\}$ indicate the left and right ears, $n = 1, \ldots, N$ is the time index, $b = 1, \ldots, N_{\text{B}}$ is the band index, $g_b[n]$ is the corresponding gammatone filter and $h^{\boldsymbol{\varphi}}_{\zeta}[n]$ is the binaural signal in a normalized scale with sound direction $\boldsymbol{\varphi}$ (i.e. a pair of head-related impulse responses or their convolution with a source signal).

We thus defined the spectral feature for the magnitude profiles (MPs) with the vector $\boldsymbol{x}_{\zeta,\text{MP}}$. This vector is the collection of root mean square amplitudes across time in decibels for each of the spectral bands for each ear:

$$\begin{gathered} \text{mp}^{\boldsymbol{\varphi}}_{\zeta,b} = 10 \log_{10}\left(\frac{1}{N}\sum_{n=1}^{N} c^{\boldsymbol{\varphi}}_{\zeta,b}[n]^2\right), \\ \boldsymbol{x}_{\zeta,\text{mp}} = \left[\text{mp}^{\boldsymbol{\varphi}}_{\zeta,1}, \ldots, \text{mp}^{\boldsymbol{\varphi}}_{\zeta,N_B}\right], \end{gathered} \tag{4}$$

where the function $c^{\boldsymbol{\varphi}}_{\zeta,b}[n]$ is defined in equation (3).

Positive gradient extraction over the frequency dimension can be computed as an alternative spectral feature since its integration increases the agreement between human localization performance and the model's predictions [15]. Therefore, we defined a second possible spectral feature based on gradient profiles (GPs) with the vector $\boldsymbol{x}_{\zeta,\text{GP}}$. It includes the gradient extraction as an additional processing step:

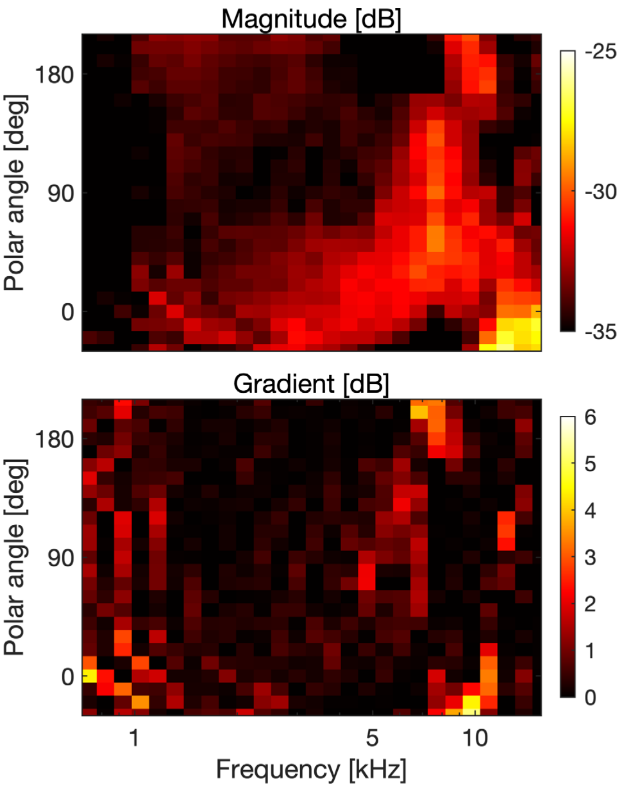

Figure 3. Monaural spectral features as a function of polar angle in the median plane. Top: Features obtained from the magnitude profiles. Bottom: Features obtained from the gradient profiles. Example for the left ear of subject NH12 [23].

$$\overline{\mathrm{gp}}^{\boldsymbol{\varphi}}_{\zeta,b} = \mathrm{mp}^{\boldsymbol{\varphi}}_{\zeta,b+1} - \mathrm{mp}^{\boldsymbol{\varphi}}_{\zeta,b},$$

$$\mathrm{gp}^{\boldsymbol{\varphi}}_{\zeta,b} = \begin{cases} \overline{\mathrm{gp}}^{\boldsymbol{\varphi}}_{\zeta,b} & \text{if } \overline{\mathrm{gp}}^{\boldsymbol{\varphi}}_{\zeta,b} \geq 0 \\ 0 & \text{otherwise}, \end{cases} \tag{5}$$

$$\boldsymbol{x}_{\zeta,\mathrm{GP}} = \left[\mathrm{gp}^{\boldsymbol{\varphi}}_{\zeta,1}, \cdots, \mathrm{gp}^{\boldsymbol{\varphi}}_{\zeta,N_B-1}\right].$$

A visualization of these monaural features is shown in Figure 3.

To demonstrate the impact of monaural spectral feature type, we analyzed the results of both variants with the corresponding feature spaces defined as follows:

$$\bar{\boldsymbol{t}}_{\mathrm{MP}} = [x_{\mathrm{itd}}, x_{\mathrm{ild}}, \boldsymbol{x}_{\mathrm{L,MP}}, \boldsymbol{x}_{\mathrm{R,MP}}],$$
$$\bar{\boldsymbol{t}}_{\mathrm{GP}} = [x_{\mathrm{itd}}, x_{\mathrm{ild}}, \boldsymbol{x}_{\mathrm{L,GP}}, \boldsymbol{x}_{\mathrm{R,GP}}]. \tag{6}$$

Limited precision in the feature extraction process leads to corruption of the features and can be modelled as additive internal noise [6]. Hence, we defined the noisy internal representation of the target features as:

$$\boldsymbol{t} = \bar{\boldsymbol{t}} + \boldsymbol{\delta},$$
$$\boldsymbol{\delta} \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{\Sigma}), \tag{7}$$

where $\boldsymbol{\Sigma}$ is the covariance matrix of the multivariate Gaussian noise. Furthermore, we assumed each spatial feature to be processed independently and thus to be also corrupted by independent noise [1]. Hence, the covariance matrix $\boldsymbol{\Sigma}$ definition is:

$$\boldsymbol{\Sigma} = \begin{bmatrix} \sigma^2_{\mathrm{itd}} & 0 & 0 \\ 0 & \sigma^2_{\mathrm{ild}} & 0 \\ 0 & 0 & \sigma^2_{\mathrm{mon}}\mathbf{I} \end{bmatrix}, \tag{8}$$

with σ^2_{itd} and σ^2_{ild} being the variances associated with the ITDs and ILDs and $\sigma^2_{\mathrm{mon}}\boldsymbol{I}$ being the covariance matrix for the monaural features where $\boldsymbol{I}$ is the identity matrix and the scalar σ_{mon} represents a constant and identical uncertainty for all frequency bands.

2.1.2 Bayesian inference

The observer infers the sound direction $\boldsymbol{\varphi}$ from the spatial features in $\boldsymbol{t}$ while taking into account potential prior beliefs about the sound direction. Within the Bayesian inference framework [5], this requires to weight the likelihood $p(\boldsymbol{t}|\boldsymbol{\varphi})$ with the prior $p(\boldsymbol{\varphi})$ to obtain the posterior distribution by means of Bayes' law as:

$$p(\boldsymbol{\varphi}|\boldsymbol{t}) \propto p(\boldsymbol{t}|\boldsymbol{\varphi})p(\boldsymbol{\varphi}). \tag{9}$$

The likelihood function was implemented by comparing $\boldsymbol{t}$ with the feature templates. The template $\boldsymbol{T}(\boldsymbol{\varphi})$ contains noiseless features of equation (1) for every sound direction $\boldsymbol{\varphi}$ [6]. While the sound direction was defined on a continuous support, our implementation sampled it over a quasi-uniform spherical grid with a spacing of 4.5° between points ($N_{\varphi} = 1500$ over the full sphere). Template features were computed from the listener-specific HRTFs. To accommodate non-uniform HRTF acquisition grids, we performed spatial interpolation based on spherical harmonics with order $N_{\mathrm{SH}} = 15$, followed by Tikhonov regularization [41].

Since the templates were constructed without noise, there exists a one-to-one mapping between direction and template features. This allowed us to write the likelihood function for each point of the direction grid as:

$$p(\boldsymbol{t}|\boldsymbol{\varphi}) = p(\boldsymbol{t}|\,\boldsymbol{T}(\boldsymbol{\varphi})) = \mathcal{N}(\boldsymbol{t}|\,\boldsymbol{T}(\boldsymbol{\varphi}), \boldsymbol{\Sigma}), \tag{10}$$

where $\boldsymbol{\Sigma}$ represents the learned precision of the auditory system (i.e. the sensory uncertainty $\boldsymbol{d}$ reported in Eq. (7)). Finally, we interpreted the a-priori probability $p(\boldsymbol{\varphi})$ to reflect long-term expectations of listeners where prior probabilities were modelled as uniformly distributed along the horizontal dimension but centered towards the horizon as [8]. In particular, we extended the results from Ege *et al.* for sources positioned in the front and as well as back positions with:

$$p(\boldsymbol{\varphi}) \propto \exp\left(-\frac{\epsilon^2}{2\sigma^2_{P,\epsilon}}\right), \tag{11}$$

with ϵ denoting the elevation angle of $\boldsymbol{\varphi}$ and $\sigma^2_{P,\epsilon}$ the variance of the prior distribution [8]. For simplicity, the prior definition was based on the spherical coordinate system.

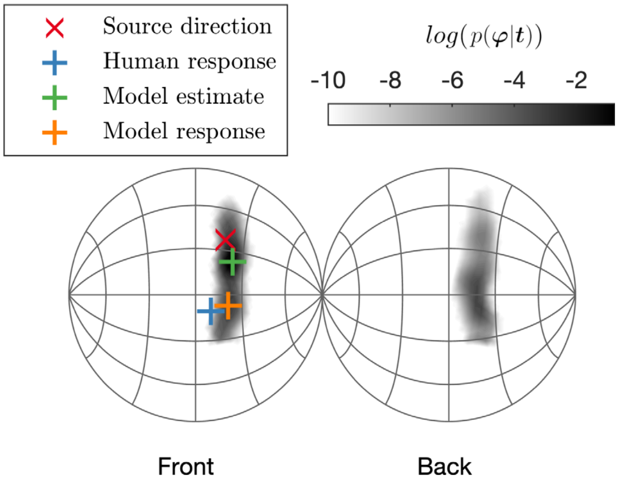

Figure 4. Example of the model estimating the direction of a broadband sound source. Red: Actual direction of the sound source. The grayscale represents the posterior probability distribution $p(\boldsymbol{\varphi}|\boldsymbol{t})$, shown, in order to increase the readability, on a logarithmic scale. Green: Direction inferred by the Bayesian-inference stage (without the response stage). Orange: Direction inferred by the model (with the response stage). Blue: Actual response of the subject.

Importantly, the origin of that prior is currently unknown and its implications are discussed in Section 4.

According to equation (9), a posterior spatial probability distribution was computed for every sound by optimally combining sensory evidence with prior knowledge [28]. As shown in Figure 4, the most probable direction of the source $\boldsymbol{\varphi}$ was then selected as the maximum a-posteriori (MAP) estimate:

$$\hat{\boldsymbol{\varphi}} = \arg\max_{\boldsymbol{\varphi}} p(\boldsymbol{t}|\boldsymbol{T}(\boldsymbol{\varphi}))p(\boldsymbol{\varphi}). \qquad (12)$$

2.1.3 Response stage

After the inference of the sound direction, experiments usually require the listener to provide a motor response (e.g. manual pointing). To account for the uncertainty introduced by such responses, we incorporated post-decision noise in the model's response stage. Following the approach from previous work [15], we blurred the location estimate by an additive, direction-independent (i.e., isotropic) Gaussian noise:

$$\hat{\boldsymbol{\varphi}}_r = \hat{\boldsymbol{\varphi}} + \boldsymbol{m}, \qquad (13)$$

where $\boldsymbol{m} \sim v\mathrm{MF}(0, \kappa_m)$ is a von-Mises–Fisher distribution with zero mean and concentration parameter κ_m. The concentration parameter κ_m can be interpreted as a standard deviation $\sigma_m = \kappa_m^{-2} \cdot 180\pi^{-1}$ [deg]. The contribution of the response noise is visible in Figure 4, where the final estimate was scattered independently of the spatial information provided by the a-posteriori distribution. With equation (13), the model outputs the response of the estimated sound source direction.

2.2 Parameter estimation

The model includes the following free parameters: σ_{ild}, σ_{mon} (amount of noise per feature; σ_{itd} was fixed to 0.569 as in [6]), $\sigma_{P,\epsilon}$ (directional prior), and σ_m (amount of response noise). Because of the model's structure, these parameters jointly contribute to the prediction of performance in both lateral and polar dimensions. To roughly account for listener-specific differences in localization performance [2], the parameters were fitted to match individual listener performance.

As for the objective fitting function, we selected a set of performance metrics widely used in the analysis of behavioral localization experiments [22, 23, 42], for a summary see [43]. A commonly used set of metrics contains the quadrant error rate (QE, i.e., frequency of polar errors larger than 90°), local polar errors (PE, i.e., root mean square error in the polar dimension that are smaller than 90°, limited to lateral angles in the range of ±30°), and lateral errors (LE, i.e., root mean square error in the lateral dimension) [22]. We accounted for the inherent stochasticity of the model estimations by averaging the simulated performance metrics over 300 repetitions of the $N_{\varphi} = 1550$ directions in the HRTF dataset (i.e., Monte–Carlo approximation with 465,000 model simulations). Model parameters were jointly adjusted in an iterative procedure (see below) until the relative residual between the actual performance metric E_a and the predicted performance metric E_p was minimized below a metric-specific threshold τ_E, i.e.,

$$|E_a - E_p|\frac{1}{E_a} < \tau_E. \qquad (14)$$

We set the thresholds to $\tau_{\mathrm{LE}} = 0.1$, $\tau_{\mathrm{PE}} = 0.15$, and $\tau_{\mathrm{QE}} = 0.2$ because these values were feasible for all subjects. In addition, the QE was transformed with the rationalized arcsine function to handle small and large values adequately [44].

We ran the estimation procedure separately for each feature space in equation (6) and each listener. First, initial values of the parameters were derived from previous literature: the variance of the prior distribution was set to $\sigma_{P,\epsilon} = 11.5°$ as in [8]. The interaural feature noise was set to $\sigma_{\mathrm{ild}} = 1$ dB, reflecting the range of ILD thresholds for pure tones [45]. The starting value for the monaural feature noise was set to $\sigma_{\mathrm{mon}} = 3.5$ dB as in [6]. The response noise standard deviation was set to $\sigma_m = 17°$ as the sensorimotor scatter found in [15]. Second, in an iterative procedure, σ_m was optimized to minimize the residual error relative to the PE metric and, similarly, σ_{mon} was adjusted to match the QE metric. Then, σ_{ild} was decreased to reach the LE metric. These steps were reiterated until the residual errors between actual and simulated metrics was less than the respective threshold. This procedure limited the σ_{m} to the interval [5°, 20°] and used a step-size of 0.1°, σ_{mon} was defined in the interval [0.5, 10] dB with a step-size of 0.05 dB; σ_{ild} was defined in the interval of [0.5, 2] dB with a step-size of 0.5 dB. If the procedure did not converge, we decreased $\sigma_{P,\epsilon}$ by 0.5° and reattempted the parameter optimization procedure.

3 Results

We first report the quality of model fits to the calibration data itself [23] in Section 3.1. Then, Section 3.2 quantitatively evaluates the simulated performances of our two model variants and of two previously proposed models against data from two additional sound localization experiments.

3.1 Parameter fits

We run the parameter estimation procedure for both model variants, based on either $\bar{\boldsymbol{t}}_{\mathrm{MP}}$ or $\bar{\boldsymbol{t}}_{\mathrm{GP}}$, and for five individuals tested in a previous study [23]. In that experiment, naive listeners were asked to localize broadband noise bursts of 500 ms duration presented from various directions on the sphere via binaural rendering through headphones based on listener-specific directional transfer functions. The subjects were wearing a head-mounted display and were asked to orient the pointer in their right hand to the perceived sound-source direction. The fitting procedure converged for both models and all subjects. Notably, subject NH15 required to reduce the step size of σ_m to 0.1° to meet the convergence criteria. Table 1 reports the estimated parameters σ_m, $\sigma_{P,\epsilon}$, σ_{mon} and σ_{ild} for every listener. The amount of response noise was similar for both model types. Table 2 contrasts the predicted performance metrics with the actual ones, averaged across listeners.

More in detail, Figure 5 compares predicted localization performance to the actual performance of subjects estimating the direction of a noise burst for different spherical segments [23]. The predicted LEs and PEs, both as functions of the actual lateral and polar angles, respectively, were in good agreement with those from the actual experiment. Instead, the simulated QE metric failed to mimic the front back asymmetries present in four subjects. Finally, only small differences were observed between the two feature spaces $\bar{\boldsymbol{t}}_{\mathrm{MP}}$ and $\bar{\boldsymbol{t}}_{\mathrm{GP}}$.

Contribution of model stages

Figure 6 illustrates the effects of different model stages on target-specific predictions. The example shows direction estimations from subject NH16 localizing broadband noise bursts [23] and the corresponding predictions of the model based on $\bar{\boldsymbol{t}}_{\mathrm{GP}}$ with different configurations of priors and response noise: without both (a), with priors only (b), and with both (c). While adding response noise scattered the estimated directions equally across spatial dimensions (compare c to b), including the spatial prior only affected the polar dimension (compare b to a). As observed in the actual responses, the prior caused more of the simulated estimations to be biased towards the horizon (0° and 180°).

In order to quantify the effect of introducing the spatial prior in the polar dimension, we computed the polar gain as a measure of accuracy [13] for both simulated and the actual responses. This metric relies on two regressions performed on the baseline condition, separating between targets in the front and back. The linear fits for the baseline condition were defined as:

$$\phi_e = g_\phi \cdot \phi_a + b_\phi \qquad (15)$$

with ϕ_e being the estimated polar angles and ϕ_a being the actual polar angles. The parameters were the localization bias b_ϕ in degrees, which is typically very small, and the dimensionless localization gain g_ϕ, which can be seen as a measure of accuracy [8, 13]. The regression fits only incorporated ϕ_e that deviate from the regression line by less than 40°. Since that definition of outliers depended on the regression parameters, this procedure was initialized with $b_\phi = 0°$ and $g_\phi = 1$ and re-iterated until convergence. In our analysis, only the frontal positions were considered. The polar gain of the actual responses, averaged over subjects, was 0.50, indicating that our subjects showed a localization error increasing with the angular distance to the horizontal plane. For the models without the prior, the predicted polar gain was 1.00 (Fig. 6a). The polar gain obtained by the model including the prior was 0.62 (Figs. 6b and 6c) showing a better correspondence to the actual polar gain. Hence, the introduction of the prior belief improved the agreement with the actual localization responses by biasing them towards the horizon.

Table 1. Fitted parameters for both model variants where the monaural spectral features were either magnitude profiles (MPs) or gradient profiles (GPs). Subjects' performance from [23].

Variant	Subject	$\sigma_{P,\epsilon}$ [deg]	σ_{ild} [dB]	σ_{mon} [deg]	σ_m [deg]
MP	NH12	11.50	0.50	3.40	8.50
	NH15	10.00	0.50	3.20	14.27
	NH16	11.50	1.00	3.60	11.00
	NH17	11.50	0.50	4.10	14.30
	NH18	11.50	1.00	6.50	14.00
GP	NH12	11.50	0.50	1.10	8.50
	NH15	11.00	0.50	1.25	14.30
	NH16	11.50	1.00	1.25	11.50
	NH17	11.50	1.00	1.60	14.00
	NH18	11.50	1.00	2.10	15.00

Table 2. Predicted performance metrics averaged across all subjects and directions (±1 standard deviation across subjects) for both model variants. Actual data from [33].

Metric	Actual	Predicted	
		MP	GP
LE [deg]	12.25 ± 2.43	12.97 ± 2.50	13.18 ± 2.66
PE [deg]	32.73 ± 3.44	31.20 ± 4.04	29.78 ± 4.01
QE [%]	7.83 ± 7.11	8.32 ± 5.75	9.80 ± 5.23

3.2 Model evaluation

The performance evaluation was done at the group-level. For our model, we used the five calibrated parameter sets with templates $\boldsymbol{T}(\boldsymbol{\varphi})$ based on the individuals' HRTFs as "digital observers". Group-level results of these digital observers were then evaluated for two psychoacoustic experiments with acoustic stimuli as input that differed

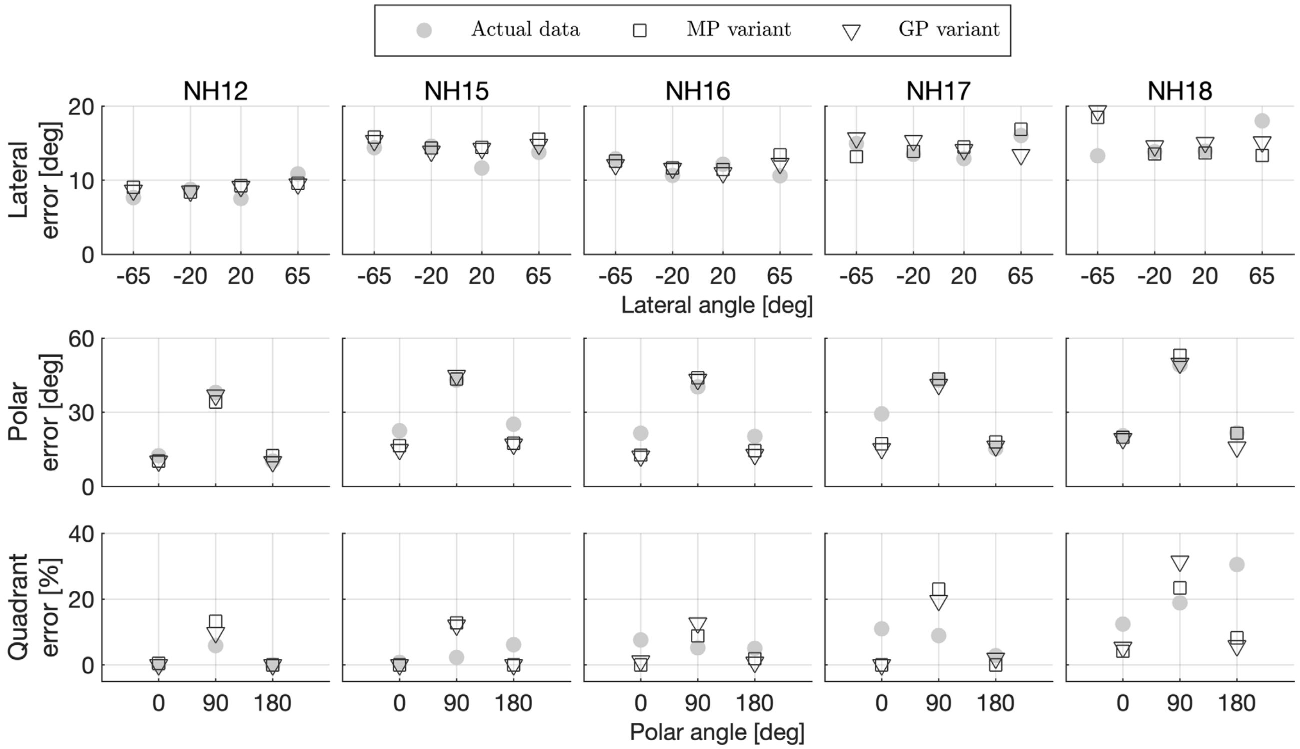

Figure 5. Sound-localization performance as function of the direction of a broadband sound source. Open symbols: Predictions obtained by the two model's variants based on either spectral magnitude profiles (MPs) or gradient profiles (GPs). Filled grey symbol: Actual data from [23]. Top row: Lateral error, calculated for all targets with lateral angles of −65° ± 25°, −20° ± 20°, 20° ± 20°, and 65° ± 25°. Center and bottom rows: Polar error and quadrant error rates, respectively, calculated for all median-plane (±30°) targets with polar angles of 0° ± 30°, 90° ± 60°, and 180° ± 30°.

from the baseline condition with a flat source spectrum and individual HRTFs.

In addition, we compared our results with the ones of two previously published models. The first one, described by Reijniers *et al.* [6], is probabilistic and able to jointly estimate the lateral and polar dimensions similar to the model described in this work. Reijniers' model deviates from ours because it relies on a different feature extraction stage, uses a uniform spatial prior distribution, does not include response noise (Eq. (13)), and does not fit individualized parameters. The second model, described by Baumgartner *et al.* [15], estimates sound positions only in the polar dimension. Nevertheless, it shares a similar processing pipeline with our model since both consider a perceptually relevant feature extraction stage, response noise, and individualized parameters. The main differences with our model are the incorporation of a directional prior, the integration of the lateral dimension, and a different method to compute the sensory evidence. Notably, this previous work implemented the template comparison procedure with the l^1-norm, which is substantially different from our likelihood function in equation (10). At the moment, the Baumgartner *et al.* model is commonly used by the scientific community interested in predictions of the elevation perception based on monaural spectral features (e.g., [46, 47]). We refer to these two models as `reijniers2014` and `baumgartner2014`, respectively.

3.2.1 Effects of non-individual HRTFs

In first evaluation, sounds were spatialized using non-individualized HRTFs [22]. Originally, eleven listeners localized Gaussian white noise bursts with a duration of 250 ms and sound directions were randomly sampled from the full sphere. Subjects were asked to estimate the direction of sounds that were spatialized using their own HRTFs in addition to sounds that were spatialized using up to four HRTFs from other subjects (21 cases in total). With the aim to reproduce these results, we had our pool of five digital listeners localize sounds from all available directions that were spatialized with their own individual HRTFs (*Own*) as well as sounds that were spatialized with HRTFs from the other four individuals (*Other*). We thus considered all inter-listener HRTF combinations for the non-individual condition.

Figure 7 summarizes the results obtained for localization experiments with own and other HRTFs. In the *Own* condition, there was a small deviation between the actual results from [22] and our model predictions. This mismatch reflects the fact that the digital observers represented a different pool of subjects (taken from [23]) tested on a slightly different experimental protocol and setup. Differences in performance metrics were small between the two feature spaces, as already reported during parameter fitting. Predictions from the `baumgartner2014` model are only possible for

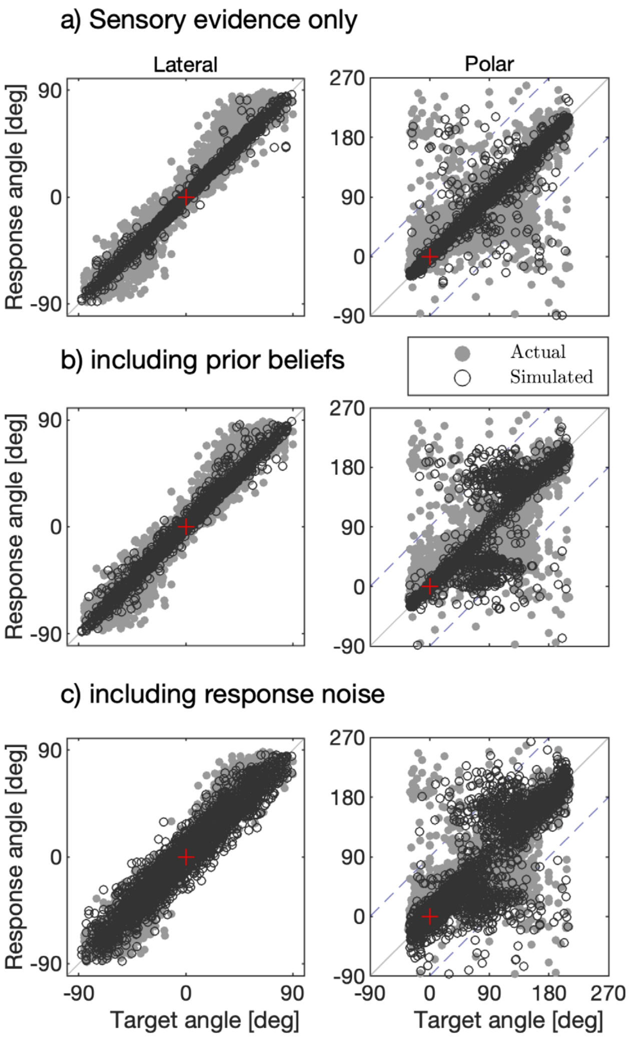

Figure 6. Effects of likelihood, prior, and response noise on predicted response patterns as a result of modeling the directional localization of broadband noise bursts. (a) Likelihood obtained by sensory evidence (i.e., no spatial prior and no response noise). (b) Bayesian inference (with the spatial prior but no response noise). (c) Full model (with prior and response noise). Gray: actual data of NH16 from [23]. Black: estimation obtained by the model considering spectral gradient profiles (GPs). Red cross: frontal position. Blue dashed lines separate regions of front-back confusions.

the polar dimension. Instead, the model reijniers2014 predicted too small errors, as also observed in previous simulations employing this model [27, 48].

In the *Other* condition, both of our model variants predicted a smaller degradation for the lateral dimension as compared to the actual data. The lateral errors predicted by reijniers2014 increased moderately but remained too small in comparison to the actual data. In the polar dimension, both model variants resulted in increased PEs and QEs, but the amount of increase was larger and more similar to the actual data for the variant equipped with gradients profiles, especially with respect to QE. As expected, the predictions from baumgartner2014 were similar to the model based on spectral gradients, given the similar method of extracting the monaural spectral features. The simulation results for reijniers2014 agreed with the super-human performance described in [27].

3.2.2 Effects of rippled-spectrum sources

The second evaluation tested the effect of spectral modulation of sound sources on directional localization in the polar dimension [33]. In that study, localization performance was probed by using noises in the frequency band [1, 16] kHz, which spectral shape were distorted with a sinusoidal modulation in the log-magnitude domain. The conditions considered different ripple depths, defined as the peak-to-peak difference of the log-spectral magnitude, and ripple densities, defined as the sinusoidal period along the logarithmic frequency scale. The actual experiment tested six trained subjects in a dark, anechoic chamber listening to the stimuli via loudspeakers. The sounds lasted 250 ms and were positioned between lateral angles of ±30° and polar angles of either 0 ± 60° for the front or 180 ± 60° for the back. A "baseline" condition included a broadband noise without any spectral modulation (ripple depth of 0 dB). To quantify the localization performance, we used the polar error rate (PER) as they defined [33]. For every condition, two baseline regressions were computed as in Section 3.1 allowing us to quantify the PER as the ratio of actual responses deviating by more than 45° from the predicted values of the baseline regression.

Figure 8 shows the results of testing the fitted models with rippled spectra. In the baseline condition, our model exhibited similar performances to those obtained in the actual experiment, whereas baumgartner2014 underestimated the baseline performance for this particular error metric. In the ripple conditions, actual listeners showed the poorest performance for densities around one ripple per octave and a systematic increase in error rate with increasing ripple depth. The model variant based on gradient profiles predicted these effects well, similar to the predictions from baumgartner2014. In contrast, both reijniers2014 and the variant based on magnitude profiles were not able to reflect the effects of ripple density and depth as present in the actual data. Hence, the positive gradient extraction appears to be a crucial processing step for predicting sagittal-plane localization of sources with a non-flat spectrum.

4 Discussion

The proposed functional model aims at reproducing listeners' performances when inferring the direction of a broadband sound source. The model formulation relies on Bayesian inference [28] as it integrates the sensory evidence for spatial directions obtained by combining binaural and monaural features [13] with a spatial prior [8]. Our approach considers uncertainties about the various sensory features, as in [6], in addition to the noise introduced by pointing responses [15]. These components enabled us to

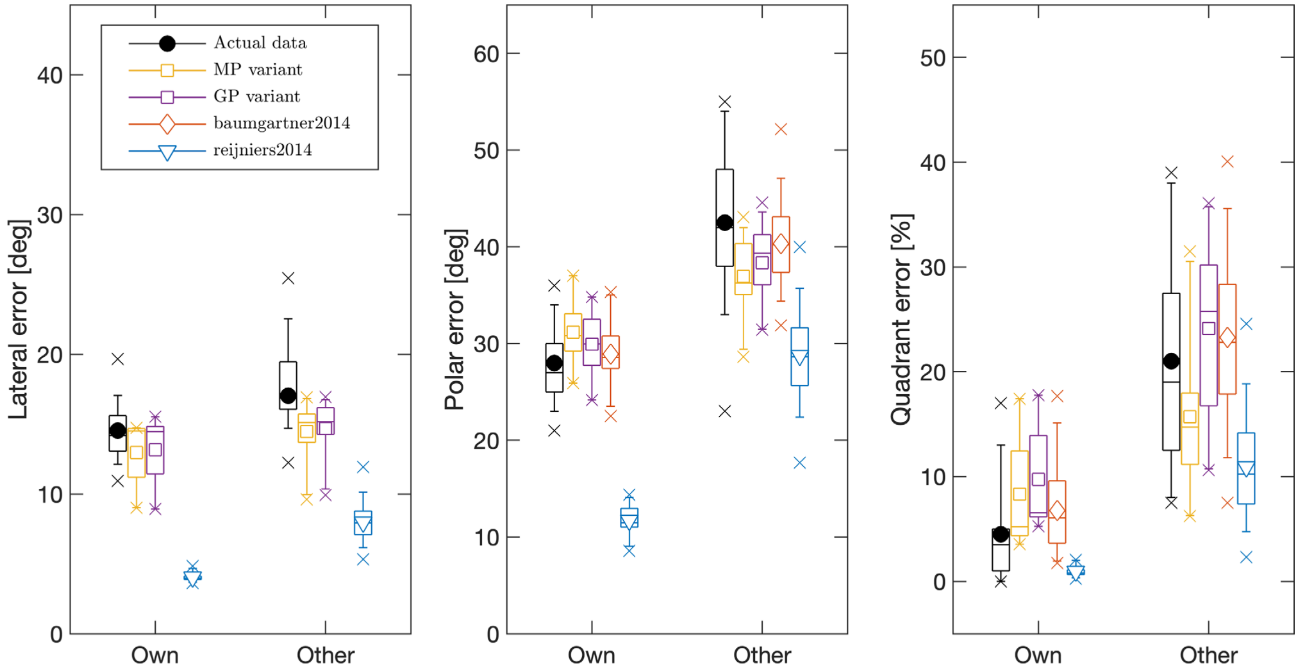

Figure 7. Localization performance with individual (*Own*) and non-individual (*Other*) HRTFs. Actual data from [22] (data_middlebrooks1999). Model predictions for two model variants: spectral magnitude profiles (MPs) and spectral gradient profiles (GPs). As references, predictions by the models reijniers2014 [6] and baumgartner2014 [15] are shown. Note that baumgartner2014 does not predict the lateral error.

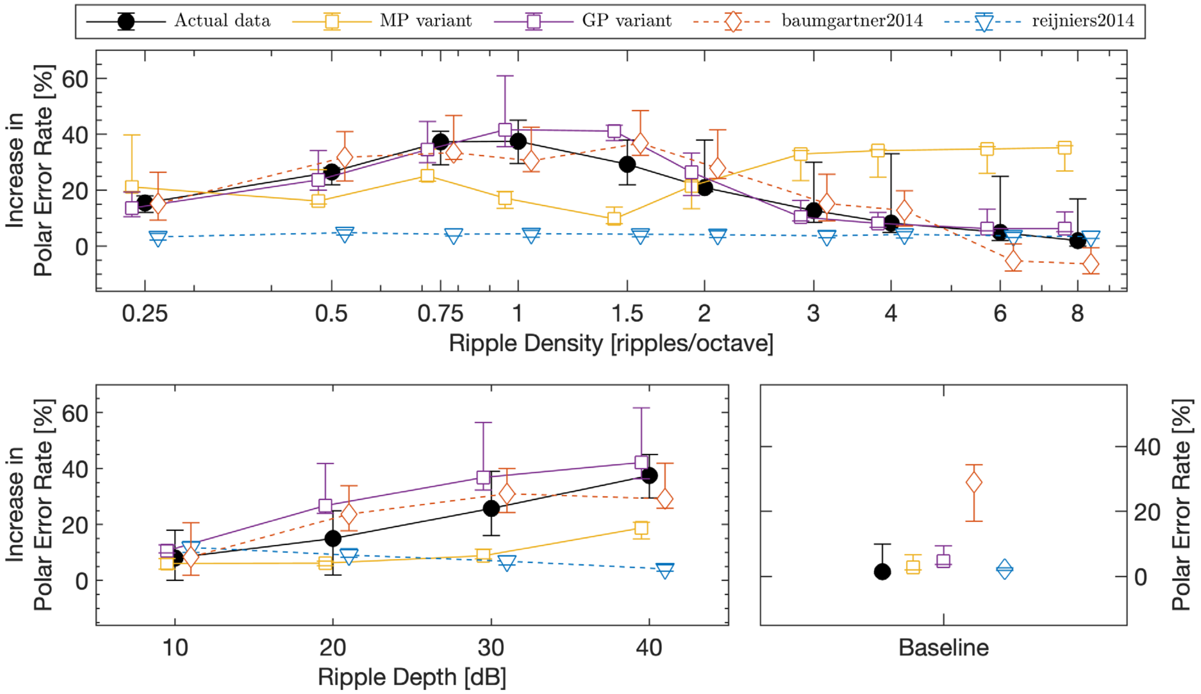

Figure 8. Effect of spectral ripples in the source spectrum on sound localization performance in the median plane. Right-most bottom panel: localization error rates obtained without spectral ripples serving as reference. Top and bottom left panels: Differences to the reference condition shown in the right-most bottom panel. In addition to predictions from the two model variants (MP and GP), predictions from reijniers2014 [6] and baumgartner2014 [15] as well as actual data from [33] (data_macpherson2003) as shown. Note that ripple densities were varied at a fixed ripple depth of 40 dB and ripple depths were varied at a fixed ripple density of one ripple per octave.

successfully match model predictions with the actual subject's performance by means of overall performance metrics (LE, PE, and QE) for five subjects (see Tab. 2) within spatially restricted areas (Fig. 5). With the inclusion of a spatial prior, the model was able to adequately explain listeners' response biases towards the horizontal plane. Compared to previous models [6, 15], our model better predicted the group-level effects of non-individualized HRTFs and rippled source spectra, yet only when selecting positive spectral gradient profiles as monaural spectral features.

We evaluated the model in scenarios where the subject was spatially static and listening to a broadband and spatially static sound source in an acoustic-free field. In this scenario, we simplified the representation of interaural features to broadband ITDs and ILDs [18, 31]. Extensions to the features are required when applying our model to more complex sounds, such as music or speech (e.g., [49]). Also, modeling sound localization in a multi-talker environment requires a frequency-dependent extraction of the interaural cues by considering temporal fine-structure within a restricted frequency range or temporal envelope disparities [32]. Similarly, modeling the localisation of sound sources placed in reverberant spaces necessitates more complex feature-specific cochlear processing to account for phenomena like the precedence effect [50, 51]. Still, our model structure is open to integrating such extensions in the future. In its present form, the model is ready to assess the degree of HRTF personalization by comparing the predicted sound-localization performance obtained with one HRTF set against a set of listener-specific HRTFs [52].

In our model, the MAP decision rule selects the most probable source position posterior distribution. We preferred this estimator over the mean estimator to adequately deal with multiple modes of the posterior distribution generated by front-back confusions along sagittal planes. On the other hand, the MAP estimator disproportionately biases direction estimates towards the prior's mode, at least under conditions of high sensory uncertainty. One of many possible alternative estimators that may better describe the stochastic human localization responses is the posterior sampling [8]: the model samples the perceptual estimate from the full posterior distribution. Although often considered suboptimal, this estimator would allow the observer to adapt to novel environmental statistics [53]. However, a different estimator might affect the fitted sensory and motor noise parameters. Therefore, comparative evaluations of different estimators would require a more robust fitting procedure, which is outside the current study's scope.

The model incorporates several non-acoustic components because they are crucial in explaining human performances [2, 54]. Extending the `reijniers2014` model [6] by incorporating a spatial prior and response scatter appeared vital to explain listeners' response patterns. Without these components, fitting the model to the polar performance metrics was unfeasible [27]. First, response noise allowed us to control the response precision locally (LE and PE) while leaving the global errors (QE) unaffected. The global errors depend predominantly on the variance of noise added to the monaural features. Second, the spatial prior shapes the response patterns by introducing a bias towards the horizon [42]. As shown in Figure 6, the prior contributed to the polar component of the simulated responses, which clustered around the eye-level direction. The polar gain measure generated additional evidence, as reported in Section 3.1, where integrating the prior beliefs led to better matching the performances in the vertical dimension. We extrapolated the spatial prior's formulation from [8] by assuming a symmetric prior distribution between front and back positions. Discrepancies observed between actual and predicted global errors (Fig. 5) indicate that this assumption was likely incorrect and points towards a front-back asymmetric prior instead. Nonetheless, we can only speculate about the reasons behind such a long-term prior in spatial hearing. It might reflect the spatial distribution of sound sources during everyday exposure [55], it may stem from an evolutionary emphasis on high relevance auditory signals [4], or it could be related to the centre of gaze as observed in barn owls [56], although the processing underlying the spatial inference mechanism might be different in mammals [3].

While the model only considers spatially static listening scenarios, it sets a foundation for future work on predicting sound-localization behavior in realistic environments. For example, modeling the environment's dynamics as a chain of consecutive stationary events is promising (e.g., [7]). Sequential update of listener's beliefs by considering the posterior as the next prior appears to be a natural mechanism under the Bayesian inference scheme [28]. Our model is a well-suited basis for such investigations. A rich set of modulators might influence the mechanism of spatial hearing, and the model's prior belief is the entry point to account for many of those like accumulation to track source statistics [20, 57], visual influences on auditory spatial perception [58], or auditory attention to segregate sources [59]. Selective temporal integration appears critical when dealing with the spatial information of many natural sources in realistic scenarios. Integrating recent findings on interaural feature extraction might solve this aspect [60]. To this end, the model must account for the dynamic interaction between the listener and the acoustic field. These extensions will potentially enable the model to account for subject movements [9] and simultaneous tracking of source movements [61] while extracting spatial information from echoic scenarios [62].

5 Conclusions

We proposed a computational auditory model for the perceptual estimation of the direction of a broadband sound source based on Bayesian inference. From a binaural input, the model estimates the sound direction by combining spatial prior beliefs with sensory evidence composed of auditory features. The model parameters are interpretable and related to sensory noise, prior uncertainty, and response noise. Having fitted the parameters to match subject-specific performance in a baseline condition, we accurately predicted the localization performance observed for test conditions with non-individualized HRTFs and spectrally-modulated source spectra. Regarding spectral monaural feature extraction, the model variant based on the spectral gradient profiles performed best.

The proposed model is useful in assessing the perceptual validity of HRTFs. However, the model's domain is currently limited to spatially static conditions, but it provides a good basis for future extensions to spatially dynamic situations, spectrally dynamic signals like speech and music, and reverberant environments.

Conflict of interest

The authors declare no conflict of interest.

Funding information

This work was supported by the European Union (EU) within the project SONICOM (grant number: 101017743, RIA action of Horizon 2020, to P.M.) and the Austrian Science Fund (FWF) within the project Dynamates (grant number: ZK66, to R.Bau.).

Acknowledgments

We thank Jonas Reijniers for providing the initial implementation of his model. We are also grateful to Günther Koliander and Clara Hollomey for helping with the mathematical formulation of some model aspects.

Data availability statement

The implementation of the model presented in this manuscript is available in the Auditory Modeling Toolbox (AMT 1.2) as `barumerli2023` [63]. Data from [23] are also available in the AMT 1.2 as `data_majdak2010`. Individual HRTF datasets of these subjects are publicly available within the ARI database at http://sofacoustics.org/data/database/ari/. Moreover, the implementations of the model `baumgartner2014` presented in [15], the model `reijniers2014` from [6] and the data for the model evaluation procedure are available as `data_middlebrooks1999` and `data_macpherson2003` in the AMT 1.2.

Additional toolboxes were selected for the model implementation. To provide quasi-uniform sphere sampling the model relied on https://github.com/AntonSemechko/S2-Sampling-Toolbox. For the implementation of the von Mises-Fisher distribution, we used https://github.com/TerdikGyorgy/3D-Simulation-Visualization/.

References

1. K. van der Heijden, J.P. Rauschecker, B. de Gelder, E. Formisano: Cortical mechanisms of spatial hearing. Nature Reviews Neuroscience 20, 10 (2019) 609–623.
2. P. Majdak, R. Baumgartner, B. Laback: Acoustic and non-acoustic factors in modeling listener-specific performance of sagittal-plane sound localization. Frontiers in Psychology 5 (2014). https://doi.org/10.3389/fpsyg.2014.00319
3. B. Grothe, M. Pecka, D. McAlpine: Mechanisms of sound localization in mammals. Physiological Reviews 90, 3 (2010) 983–1012.
4. M. Pecka, C. Leibold, B. Grothe: Biological aspects of perceptual space formation. The Technology of Binaural Understanding. Springer (2020) 151–171.
5. W.J. Ma: Organizing probabilistic models of perception. Trends in Cognitive Sciences 16, 10 (2012) 511–518.
6. J. Reijniers, D. Vanderelst, C. Jin, S. Carlile, H. Peremans: An ideal-observer model of human sound localization. Biological Cybernetics 108, 2 (2014) 169–181.
7. H. Kayser, V. Hohmann, S.D. Ewert, B. Kollmeier, J. Anemüller: Robust auditory localization using probabilistic inference and coherence-based weighting of interaural cues. The Journal of the Acoustical Society of America 138, 5 (2015) 2635–2648.
8. R. Ege, A.J. van Opstal, M.M. van Wanrooij: Accuracy-precision trade-off in human sound localisation. Scientific Reports 8, 1 (2018) 16399.
9. G. McLachlan, P. Majdak, J. Reijniers, H. Peremans: Towards modelling active sound localisation based on bayesian inference in a static environment. Acta Acustica 5 (2021) 45.
10. H. Møller, M.F. Sørensen, D. Hammershøi, C.B. Jensen: Head-related transfer functions of human subjects. Journal of the Audio Engineering Society 43, 5 (1995) 300–321.
11. J.C. Middlebrooks: Narrow-band sound localization related to external ear acoustics. The Journal of the Acoustical Society of America 92, 5 (1992) 2607–2624.
12. P. Zakarauskas, M.S. Cynader: A computational theory of spectral cue localization. The Journal of the Acoustical Society of America 94, 3 (1993) 1323–1331.
13. P.M. Hofman, A.J. van Opstal: Spectro-temporal factors in two-dimensional human sound localization. The Journal of the Acoustical Society of America 103, 5 (1998) 2634–2648.
14. E.H.A. Langendijk, A.W. Bronkhorst: Contribution of spectral cues to human sound localization. The Journal of the Acoustical Society of America 112, 4 (2002) 1583–1596.
15. R. Baumgartner, P. Majdak, B. Laback: Modeling sound-source localization in sagittal planes for human listeners. The Journal of the Acoustical Society of America 136, 2 (2014) 791–802.
16. R. Baumgartner, P. Majdak, B. Laback: Modeling the effects of sensorineural hearing loss on sound localization in the median plane. Trends in Hearing 20 (2016) 2331216516662003.
17. A.J. Van Opstal, J. Vliegen, T. van Esch: Reconstructing spectral cues for sound localization from responses to rippled noise stimuli. PLoS One 12, 3 (2017) 1–29.
18. J.C. Middlebrooks: Sound localization. Handbook of Clinical Neurology, Vol. 129, Elsevier (2015) 99–116.
19. M.M. van Wanrooij, A. John van Opstal: Relearning sound localization with a new ear. Journal of Neuroscience 25, 22 (2005) 5413–5424.
20. M. Morimoto, H. Aokata: Localization cues of sound sources in the upper hemisphere. Journal of the Acoustical Society of Japan (E) 5, 3 (1984) 165–173.
21. K. Pollack, W. Kreuzer, P. Majdak: Perspective chapter: Modern acquisition of personalised head-related transfer functions – an overview, in: B.F.G. Katz, Piotr Majdak (Eds.), Advances in fundamental and applied research on spatial audio, Rijeka: IntechOpen, 2022.
22. J.C. Middlebrooks: Virtual localization improved by scaling nonindividualized external-ear transfer functions in frequency. The Journal of the Acoustical Society of America 106, 3 (1999) 1493–1510.

23. P. Majdak, M.J. Goupell, B. Laback: 3-D localization of virtual sound sources: Effects of visual environment, pointing method, and training. Attention, Perception, & Psychophysics 72, 2 (2010) 454–469.
24. D.P. Kumpik, O. Kacelnik, A.J. King: Adaptive reweighting of auditory localization cues in response to chronic unilateral earplugging in humans. Journal of Neuroscience 30, 14 (2010) 4883–4894.
25. F.L. Wightman, D.J. Kistler: Monaural sound localization revisited. The Journal of the Acoustical Society of America 101, 2 (1997) 1050–1063.
26. J.O. Stevenson-Hoare, T.C.A. Freeman, J.F. Culling: The pinna enhances angular discrimination in the frontal hemifield. The Journal of the Acoustical Society of America 152, 4 (2022) 2140–2149.
27. R. Barumerli, P. Majdak, R. Baumgartner, M. Geronazzo, F. Avanzini: Evaluation of a human sound localization model based on Bayesian inference, in Forum Acusticum, Lyon, France, December (2020) 1919–1923.
28. W.J. Ma: Bayesian decision models: A Primer. Neuron 104, 1 (2019) 164–175.
29. R. Ege, A.J. van Opstal, M.M. van Wanrooij: Perceived target range shapes human sound-localization behavior. eneuro 6(2) (2019) ENEURO.0111–18.2019.
30. K. Krishnamurthy, M.R. Nassar, S. Sarode, J.I. Gold: Arousal-related adjustments of perceptual biases optimize perception in dynamic environments. Nature Human Behaviour 1, 6 (2017) 1–11.
31. A. Andreopoulou, B.F. Katz: Identification of perceptually relevant methods of inter-aural time difference estimation. The Journal of the Acoustical Society of America 142, 2 (2017) 588–598.
32. M. Dietz, S.D. Ewert, V. Hohmann: Auditory model based direction estimation of concurrent speakers from binaural signals. Speech Communication 53, 5 (2011) 592–605.
33. E.A. Macpherson, J.C. Middlebrooks: Vertical-plane sound localization probed with ripple-spectrum noise. The Journal of the Acoustical Society of America 114, 1 (2003) 430–445.
34. D.J. Kistler, F.L. Wightman: A model of head-related transfer functions based on principal components analysis and minimum-phase reconstruction. The Journal of the Acoustical Society of America 91, 3 (1992) 1637–1647.
35. J.E. Mossop, J.F. Culling: Lateralization of large interaural delays. The Journal of the Acoustical Society of America 104, 3 (1998) 1574–1579.
36. B.R. Glasberg, B.C.J. Moore: Derivation of auditory filter shapes from notched-noise data. Hearing Research 47, 1 (1990) 103–138.
37. A. Saremi, R. Beutelmann, M. Dietz, G. Ashida, J. Kretzberg, S. Verhulst: A comparative study of seven human cochlear filter models. The Journal of the Acoustical Society of America 140, 3 (2016) 1618–1634.
38. V.R. Algazi, C. Avendano, R.O. Duda: Elevation localization and head-related transfer function analysis at low frequencies. The Journal of the Acoustical Society of America 109, 3 (2001) 1110–1122.
39. J. Hebrank, D. Wright: Spectral cues used in the localization of sound sources on the median plane. The Journal of the Acoustical Society of America 56, 6 (1974) 1829–1834.
40. N. Roman, D. Wang, G.J. Brown: Speech segregation based on sound localization. The Journal of the Acoustical Society of America 114, 4 (2003) 18.
41. D.N. Zotkin, R. Duraiswami, N.A. Gumerov: Regularized HRTF fitting using spherical harmonics, in: 2009 IEEE Workshop on Applications of Signal Processing to Audio and Acoustics, IEEE, New Paltz, NY, USA, October (2009) 257–260.
42. S. Carlile, P. Leong, S. Hyams: The nature and distribution of errors in sound localization by human listeners. Hearing Research 114, 1 (1997) 179–196.
43. P. Majdak, M.J. Goupell, B. Laback: Two-dimensional sound localization in cochlear implantees. Ear and Hearing 32, 2 (2011) 198–208.
44. G.A. Studebaker: A rationalized arcsine transform. Journal of Speech, Language, and Hearing Research 28, 3 (1985) 455–462.
45. W.A. Yost, R.H. Dye: Discrimination of interaural differences of level as a function of frequency. The Journal of the Acoustical Society of America 83, 5 (1988) 1846–1851.
46. R. Barumerli, M. Geronazzo, F. Avanzini, Localization in elevation with non-individual head-related transfer functions: comparing predictions of two auditory models, in 2018 26th European Signal Processing Conference (EUSIPCO) (2018) 2539–2543. https://doi.org/10.23919/EUSIPCO.2018.8553320. ISSN: 2076-1465.
47. D. Marelli, R. Baumgartner, P. Majdak: Efficient approximation of head-related transfer functions in subbands for accurate sound localization. IEEE/ACM Transactions on Audio, Speech, and Language Processing 23, 7 (2015) 1130–1143. https://doi.org/10.1109/TASLP.2015.2425219.
48. R. Barumerli, P. Majdak, J. Reijniers, R. Baumgartner, M. Geronazzo, F. Avanzini: Predicting directional sound-localization of human listeners in both horizontal and vertical dimensions. Audio Engineering Society Convention 148 (2020).
49. V. Best, S. Carlile, C. Jin, A. van Schaik, The role of high frequencies in speech localization, The Journal of the Acoustical Society of America 118, 1 (2005) 353–363. https://doi.org/10.1121/1.1926107.
50. J. Blauert: Spatial hearing. The Psychophysics of Human Sound Localization. The MIT Press, Cambridge, MA, revised edition (1997).
51. R. Ege, A.J. van Opstal, P. Bremen, M.M. van Wanrooij: Testing the precedence effect in the median plane reveals backward spatial masking of sound. Scientific Reports 8, 1 (2018) 8670.
52. M. Geronazzo, S. Spagnol, F. Avanzini: Do we need individual head-related transfer functions for vertical localization? The case study of a spectral notch distance metric. IEEE/ACM Transactions on Audio, Speech, and Language Processing 26, 7 (2018) 1243–1256.
53. W. Gaissmaier, L.J. Schooler: The smart potential behind probability matching. Cognition 109, 3 (2008) 416–422.
54. G. Andeol, E.A. Macpherson, A.T. Sabin: Sound localization in noise and sensitivity to spectral shape. Hearing Research 304 (2013) 20–27.
55. C.V. Parise, K. Knorre, M.O. Ernst: Natural auditory scene statistics shapes human spatial hearing. Proceedings of the National Academy of Sciences 111, 16 (2014) 6104–6108.
56. B.J. Fischer, J.L. Peña: Owl's behavior and neural representation predicted by Bayesian inference. Nature Neuroscience 14, 8 (2011) 1061–1066.
57. B. Skerritt-Davis, M. Elhilali: Detecting change in stochastic sound sequences. PLOS Computational Biology 14, 5 (2018) 1–24.
58. B. Odegaard, U.R. Beierholm, J. Carpenter, L. Shams: Prior expectation of objects in space is dependent on the direction of gaze. Cognition 182 (2019) 220–226.
59. E.M. Kaya, M. Elhilali: Modelling auditory attention. Philosophical Transactions of the Royal Society B: Biological Sciences 372, 1714 (2017) 20160101.
60. M. Dietz, T. Marquardt, N.H. Salminen, D. McAlpine: Emphasis of spatial cues in the temporal fine structure during the rising segments of amplitude-modulated sounds. Pro-

ceedings of the National Academy of Sciences 110, 37 (2013) 15151–15156.

61. D.A. Hambrook, M. Ilievski, M. Mosadeghzad, M. Tata: A Bayesian computational basis for auditory selective attention using head rotation and the interaural time-difference cue. PLOS One 12, 10 (2017) e0186104.
62. D.B. Ward, E.A. Lehmann, R.C. Williamson: Particle filtering algorithms for tracking an acoustic source in a reverberant environment. IEEE Transactions on Speech and Audio Processing 11, 6 (2003) 826–836.
63. P. Majdak, C. Hollomey, R. Baumgartner: Amt 1.x: A toolbox for reproducible research in auditory modeling. Acta Acustica 6 (2022) 19.

Cite this article as: Barumerli R. Majdak P. Geronazzo M. Meijer D. Avanzini F, et al. 2023. A Bayesian model for human directional localization of broadband static sound sources. Acta Acustica, **7**, 12.