Biol Cybern (2014) 108:169–181
DOI 10.1007/s00422-014-0588-4

ORIGINAL PAPER

Biological Cybernetics

An ideal-observer model of human sound localization

J. Reijniers · D. Vanderelst · C. Jin · S. Carlile · H. Peremans

Received: 10 June 2013 / Accepted: 24 January 2014 / Published online: 26 February 2014

Abstract In recent years, a great deal of research within the field of sound localization has been aimed at finding the acoustic cues that human listeners use to localize sounds and understanding the mechanisms by which they process these cues. In this paper, we propose a complementary approach by constructing an ideal-observer model, by which we mean a model that performs optimal information processing within a Bayesian context. The model considers all available spatial information contained within the acoustic signals encoded by each ear. Parameters for the optimal Bayesian model are determined based on psychoacoustic discrimination experiments on interaural time difference and sound intensity. Without regard as to how the human auditory system actually processes information, we examine the best possible localization performance that could be achieved based only on analysis of the input information, given the constraints of the normal auditory system. We show that the model performance is generally in good agreement with the actual human localization performance, as assessed in a meta-analysis of many localization experiments (Best et al. in Principles and applications of spatial hearing, pp 14–23. World Scientific Publishing, Singapore, 2011). We believe this approach can shed new light on the optimality (or otherwise) of human sound localization, especially with regard to the level of uncertainty in the input information. Moreover, the proposed model allows one to study the relative importance of various (combinations of) acoustic cues for spatial localization and enables a prediction of which cues are most informative and therefore likely to be used by humans in various circumstances.

Keywords Human hearing · Sound localization · Ideal-observer model · Bayesian model · Head-related transfer function · Localization performance

Electronic supplementary material The online version of this article (doi:10.1007/s00422-014-0588-4) contains supplementary material, which is available to authorized users.

J. Reijniers (✉)
Biology Department, University of Antwerp, Antwerp, Belgium
e-mail: jonas.reijniers@uantwerpen.be

D. Vanderelst · H. Peremans
Department of Engineering Management,
University of Antwerp, Antwerp, Belgium

C. Jin
Department of Electrical and Information Engineering,
University of Sydney, Sydney, NSW, Australia

S. Carlile
School of Medical Sciences and Institute for Biomedical Research,
University of Sydney, Sydney, NSW, Australia

1 Introduction

The acoustic cues used by humans in locating sound sources have been extensively studied, both experimentally and theoretically (Blauert 1997). Assessing the relative contribution of different information sources is not only important to advance our understanding of human hearing but is also of practical interest, as the new generation of hearing aids has developed to the point that spatial hearing can be targeted (Best et al. 2010; Boyd et al. 2012).

Psychoacoustic experiments have revealed the importance of both interaural time difference (ITD) and interaural level difference (ILD) cues in encoding the lateral position of a sound source, i.e., the sagittal planes which contains the so-called cones of confusion (Woodworth 1938). Subsequent modeling has shown that ITD and ILD cues essentially contain redundant spatial information, defining near-coinciding cones of confusion, about sound sources in the distal region (range > 1m) of the listener's head (Blauert 1997). However, in the near field (range < 1m), combining both cues leads to decreased ambiguity about the sound source location by reducing the possible positions to 'tori of confusion' (Shinn-Cunningham et al. 2000). Psychoacoustic experiments (Musicant and Butler 1984; Wightman and Kistler 1989; Carlile and Pralong 1994; Kulkarni and Colburn 1998; Algazi et al. 2001) further showed that the spectra of the signals received at both ears carry additional information on the position of the sound source that aid in resolving this ambiguity in the binaural cues. This direction-dependent spectral filtering by the listener's body, head and ears is described by the head-related transfer function (HRTF). The combination of the HRTF, ILD and ITD cues allows the identification of the unique position (distance, azimuth and elevation) of the sound source on the tori of confusion in the near field (Brungart and Rabinowitz 1999). In the far field (> 1m), combining these cues allows the identification of the unique location (azimuth and elevation) on the cone of confusion. Apart from these primary cues, different acoustical parameters further affect localization performance (sound intensity, frequency content, reverberation, etc.), as do individual differences and experimental task differences. An excellent overview of this extensive area of research can be found in Colburn and Kulkarni (2005).

Psychoacoustic experiments are aimed at finding the acoustic cues that human listener's use, as well as understanding the mechanisms by which they transform these cues into localization estimates. In this paper, we propose a complementary approach by constructing an ideal-observer model based on a Bayesian analysis (Duda et al. 2001) of sound source localization. This model reveals the best theoretically possible localization performance that can be achieved by an optimal combination of all available information, given the limitations imposed by the precision of the human hearing system, estimated explicitly from psychoacoustic discrimination experiments. Under the assumption that subjects make optimal use of the available information, the model should predict the subjects' performance. Deviations from optimality can be explicitly modeled as biases in the listeners' decision making or may indicate specific limitations in the biological encoding or processing of this information. This approach results in three major differences with spatial hearing models developed so far.

(1) The main purpose of the proposed model is to predict how much uncertainty about source position remains, given that all *a priori* information and all spatial information contained in the received acoustic signals is optimally combined, irrespective of the particular mechanisms proposed for extracting this information from the acoustic cues. Hence, the predicted performance degradations do not depend in any way on the particular processing mechanisms active in the human observer; they are solely due to the absence of the relevant spatial information in the received acoustic signals and in the *a priori* information available to the listener. As no amount of processing can reintroduce missing information, the model defines an upper limit to human sound localization performance. While most models proposed so far are descriptive models (Middlebrooks 1992; Macpherson and Sabin 2007), normative models defining an upper performance limit have been used before. Based on an analysis of the information content of binaural ILD and ITD cues in the near field around a spherical head, it has been shown that the loci of sound source positions compatible with the observed cues can be described by the tori of confusion (Shinn-Cunningham et al. 2000). Assuming knowledge of the sound source spectrum, it is shown in Brungart and Rabinowitz (1999) that the volume of these ambiguity regions (=tori of confusion) can be further reduced by including the information contained in the HRTF cues generated by a real head.

(2) In contrast to previous modeling work, the proposed Bayesian model integrates the information carried by the acoustic signal with the *a priori* information available to the listener into a decision rule that minimizes a possibly task-dependent loss function. Indeed, apart from the spatial information contained in the acoustic signal proper, there is often additional information in the context of the localization task that can be exploited. The listener can have knowledge about the spectrum and/or intensity of the sound source which, when combined with the spatial information from the acoustic signal, allows one to improve localization performance. Similarly, the listener often has (some) *a priori* information on the location of the sound source, e.g., in a localization experiment not all sound source locations may be sampled uniformly. The listener may learn this distribution which could then affect localization performance.

(3) Humans are not necessarily ideal observers. Instead of processing all information contained in the acoustic signals received by both ears, the human brain might limit itself to processing certain features of these signals only. The proposed model allows one to investigate what spatial information would be lost, if one would consider specific cues (HRTF, ITD, ILD, etc.) or combinations thereof and their associated noise models as input for the model instead of the original acoustic signals received by both ears. Hence, although this work does not aim at modeling specific processing mechanisms directly, it can still help in understanding human sound localization by mapping out the correspondence between specific features (cues) and combinations thereof and the spatial information they contain. A similar approach of apply-

ing ideal observer models to non-optimal cues has also been proposed in Colburn (1973), Dabak and Johnson (1993) to determine neural implementations of human sound localization mechanisms. By controlling the non-optimal nature of the cues, i.e., encoding only particular features of the received acoustic signals or introducing physiologically realistic internal noise, the performance of the model can be varied until it corresponds with actual human performance for the same task. In Colburn (1973), this approach is taken to get insight into what aspects of the auditory-nerve patterns are actually used by the central structures. In Dabak and Johnson (1993), a similar approach, described as function-based modeling, is taken to derive functional realizations for specific neural systems from the optimal processor's structure. Note that these last two approaches base optimality on a maximum likelihood criterion which does not minimize any function, such as the loss function in the Bayesian approach, of the estimation error (Dabak and Johnson 1993).

In the proposed model, all spatial information is considered to be encoded by the time difference between the signals arriving at both ears and by the two monaural spectra of these signals. This spatial information is decoded using a probabilistic approach, i.e., the acoustic cues are compared to a set of stored templates corresponding with all the possible sound directions, and the posterior probability of each of the templates is assessed by inspecting the deviations from the measured acoustic cues. Deviations are to be expected since the spatial information is blurred by the limited precision of the hearing system, by environmental noise and by the uncertainty about the spectrum of the sound source. The model parameters are estimated from psychoacoustic discrimination experiments aimed at measuring basic properties of the hearing system, i.e., ITD and intensity just-noticeable differences (JND). Note that, while some of the mathematical steps of the proposed model seem similar to those from the so-called template matching models proposed to explain human sound localization (Middlebrooks 1992; Macpherson and Sabin 2007; Macpherson 1997; Langendijk and Bronkhorst 2002; Hofman and Van Opstal 2003), we would like to stress that our model, contrary to these models, does not intend to be a descriptive one. Hence, the transformations introduced during the model calculations do not need to be in accordance with what is known about human spatial hearing; they only need to be information preserving.

In this paper, we use the proposed model to study human localization performance over the full 2D sphere, when presented with a broadband sound source. We chose to concentrate on broadband transient sounds first, as they are considered to be important for sound localization: They are generated in many different contexts (speech, etc.), their spectral content is fairly predictable, and because of the large bandwidth, they are well suited to encode (spatial) information. As mentioned earlier, the localization performance can differ significantly between individuals, and therefore, to average out individual differences, model calculations are done for 100 individuals. For each individual, the HRTF is measured as described in Jin et al. (2004), and using this HRTF as input, the individual's localization performance is simulated. Pooling the simulated performance of these 100 subjects, we then compare the predicted average localization performance to the average human sound localization performance, as it was determined experimentally in a meta-analysis of localization trials that have been conducted over several years in 4 different laboratories (Best et al. 2011). Next, we study the sensitivity of the localization performance to the different model parameters and put forward some hypotheses to account for discrepancies between the performance according to our model and that obtained in the meta-analysis.

2 Model

In the model, we try to localize sounds in two dimensions, i.e., determine the lateral and the polar angle, as represented by the angular vector $\boldsymbol{\theta} = \{\theta_1, \theta_2\}$, shown in Fig. 1. The Cartesian coordinates are then given by $\{\cos\theta_1 \cos\theta_2; \sin\theta_1; \cos\theta_1 \sin\theta_2\}$, such that the angular vector has norm 1. The estimation of the distance to the sound source is outside the scope of this paper.

2.1 Acoustic information

We assume that the acoustic information $\boldsymbol{X}$ is relayed to the brain. This acoustic information consists of the following features.

The interaural time difference ITD: The first cue we take into account is the interaural time difference (ITD). The ITD as it is measured can be written as

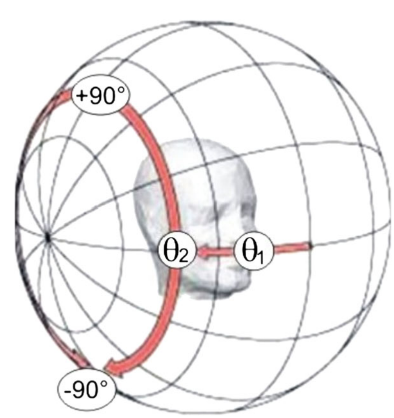

Fig. 1 The interaural polar coordinate system

$X_{\mathrm{itd}} = \mathrm{itd}(\boldsymbol{\theta}) + \delta_{\mathrm{itd}},$

with $\mathrm{itd}(\boldsymbol{\theta})$ the exact ITD corresponding to direction $\boldsymbol{\theta}$, and δ_{itd} the error on the ITD measurement, due to the limited precision of the hearing apparatus and/or environmentally induced uncertainty (e.g., air turbulence, etc).

The logmagnitude spectra in the left and the right ear: We consider the logmagnitude spectra in the left and the right ear, which are given by the frequencywise sum of the logmagnitudes of the sound source spectrum $\boldsymbol{S}$ and the HRTF for the left and right ear, respectively $\boldsymbol{H}_{\mathrm{L}}$ and $\boldsymbol{H}_{\mathrm{R}}$.

$$\boldsymbol{X}_{\mathrm{L}} = \boldsymbol{S} + \boldsymbol{H}_{\mathrm{L}}(\boldsymbol{\theta}) + \boldsymbol{\delta}_{\mathrm{L}}$$
$$\boldsymbol{X}_{\mathrm{R}} = \boldsymbol{S} + \boldsymbol{H}_{\mathrm{R}}(\boldsymbol{\theta}) + \boldsymbol{\delta}_{\mathrm{R}},$$

with the respective errors $\boldsymbol{\delta}_{\mathrm{L}}$ and $\boldsymbol{\delta}_{\mathrm{R}}$ due to the limited precision of the hearing apparatus in measuring the frequency content of the incoming sound. The symbols in bold font denote vectors with values along the frequency axis (at fixed frequencies, the so-called frequency channels, see further).

In the following, we assume $\boldsymbol{S}$ to be identical for both ears, i.e., we assume that the sound source is in the far field. The spectrum of the sound source $\boldsymbol{S}$ is not known exactly by the listener, but the listener often has some *a priori* information on the spectrum of the sound source. Indeed, in real life, the listener is confronted with a limited set of different sounds, and depending on the context and the obtained spectra, he will be able to estimate the sound source spectrum with $\hat{\boldsymbol{S}} = \boldsymbol{S} + \boldsymbol{\delta}_S$, where $\boldsymbol{\delta}_S$ is the error due to imperfect estimation. Consequently, the above equations can be rewritten as

$$\boldsymbol{X}_{\mathrm{L}} = \hat{\boldsymbol{S}} + \boldsymbol{H}_{\mathrm{L}}(\boldsymbol{\theta}) + \boldsymbol{\delta}_{\mathrm{L}} + \boldsymbol{\delta}_S \qquad (1)$$
$$\boldsymbol{X}_{\mathrm{R}} = \hat{\boldsymbol{S}} + \boldsymbol{H}_{\mathrm{R}}(\boldsymbol{\theta}) + \boldsymbol{\delta}_{\mathrm{R}} + \boldsymbol{\delta}_S.$$

The acoustic information about the sound source position is then contained in the vector

$$\boldsymbol{X} = [X_{\mathrm{itd}}, \boldsymbol{X}_{\mathrm{L}}, \boldsymbol{X}_{\mathrm{R}}],$$

where, as noted, different components of each vector correspond to different frequencies.

2.2 Projection of the acoustic information to a new basis

In the above representation, the error $\boldsymbol{\delta}_S$ on the sound spectrum is present both in $\boldsymbol{X}_{\mathrm{L}}$ and $\boldsymbol{X}_{\mathrm{R}}$. We can separate $\boldsymbol{\delta}_{\mathrm{L}}$, $\boldsymbol{\delta}_{\mathrm{R}}$ from $\boldsymbol{\delta}_S$ if we project the above vector $[\boldsymbol{X}_{\mathrm{L}}, \boldsymbol{X}_{\mathrm{R}}]$ to a new basis, which preserves the information in the original signals. The new basis reads

$$\boldsymbol{X}_- = \boldsymbol{X}_{\mathrm{L}} - \boldsymbol{X}_{\mathrm{R}} = \boldsymbol{H}_{\mathrm{L}}(\boldsymbol{\theta}) - \boldsymbol{H}_{\mathrm{R}}(\boldsymbol{\theta}) + \boldsymbol{\delta}_-$$
$$\boldsymbol{X}_+ = (\boldsymbol{X}_{\mathrm{L}} + \boldsymbol{X}_{\mathrm{R}})/2 = \hat{\boldsymbol{S}} + [\boldsymbol{H}_{\mathrm{L}}(\boldsymbol{\theta}) + \boldsymbol{H}_{\mathrm{R}}(\boldsymbol{\theta})]/2 + \boldsymbol{\delta}_+,$$

with

$$\boldsymbol{\delta}_- = \boldsymbol{\delta}_{\mathrm{L}} - \boldsymbol{\delta}_{\mathrm{R}}$$
$$\boldsymbol{\delta}_+ = \boldsymbol{\delta}_S + (\boldsymbol{\delta}_{\mathrm{L}} + \boldsymbol{\delta}_{\mathrm{R}})/2.$$

This transformation is not strictly necessary, but it will turn out to be useful to interpret and discuss the results and relate them to existing localization models, e.g., $\boldsymbol{X}_- = \boldsymbol{X}_{\mathrm{L}} - \boldsymbol{X}_{\mathrm{R}}$ corresponds to the interaural spectral difference (ISD) that is considered useful for spatial localization, in particular to estimate sound source elevation (Colburn and Kulkarni 2005). Indeed, this information is independent of knowledge of the source spectrum (no dependency on $\boldsymbol{\delta}_S$). If one does have some *a priori* knowledge about the source spectrum, there can be additional information in $\boldsymbol{X}_+$, which is essentially the average of both monaural spectra. In the new basis, the acoustic information reads

$$\boldsymbol{X} = [X_{\mathrm{itd}}, \boldsymbol{X}_-, \boldsymbol{X}_+].$$

Note that in many other models, in addition to the information in the ISD ($\boldsymbol{X}_-$), one often chooses the monaural spectrum of the proximal ear as the additional carrier of spatial information, instead of the average of both monaural spectra $\boldsymbol{X}_+$ (Colburn and Kulkarni 2005). We chose the above representation because it is symmetric, and no ad hoc bias for right or left spectrum is introduced. It is true that for lower sound intensities, system noise will mostly affect the spectral information in the contralateral ear, but we address the effect of system noise on the localization performance further on in this paper.

Note also that our analysis is intended to indicate *what* directional information can be decoded from the binaural acoustic cues. It should not be interpreted as a model for *how* the listener extracts this information to localize the sound source.

2.3 Estimating sound source position θ

2.3.1 Bayes' rule

Decoding the spatial information from the auditory input essentially comes down to the calculation of $P(\boldsymbol{\theta}|\boldsymbol{X})$, which is the probability that given the acoustic information $\boldsymbol{X}$ is measured, the sound originates from direction $\boldsymbol{\theta}$. In order to calculate $P(\boldsymbol{\theta}|\boldsymbol{X})$, we make use of Bayes' rule

$$P(\boldsymbol{\theta}|\boldsymbol{X}) = \frac{P(\boldsymbol{\theta})P(\boldsymbol{X}|\boldsymbol{\theta})}{P(\boldsymbol{X})}, \qquad (2)$$

where $P(\boldsymbol{\theta})$ describes the prior knowledge of the observer about the direction of the sound source. This prior probability need not be uniformly distributed. Indeed, it may be that based on earlier measurements, the observer is biased and favors some directions at the expense of others. Consequently, $P(\boldsymbol{\theta})$ depends to a large extent on the context.

But if we assume no prior knowledge about $\boldsymbol{\theta}$ (corresponding with an experiment where the sound source has to be localized based on a single sound presentation and where all directions are equally represented), then $P(\boldsymbol{\theta})$ is uniformly distributed, and as $P(\boldsymbol{X})$ can be inferred from the fact that $\int P(\boldsymbol{\theta}|\boldsymbol{X})\,\mathrm{d}\boldsymbol{\theta} = 1$, calculating the posterior probability $P(\boldsymbol{\theta}|\boldsymbol{X})$ is reduced to the calculation of $P(\boldsymbol{X}|\boldsymbol{\theta})$, i.e., the likelihood that a sound originating from direction $\boldsymbol{\theta}$ gives rise to the acoustic information $\boldsymbol{X}$.

2.3.2 Modeling a priori information used by the listener

In order to estimate the sound source direction, the listener does not solely rely on the acoustic information $\boldsymbol{X}$, but interprets these cues in the context of *a priori* information defining a measurement model based on knowledge of the spectral filtering $\boldsymbol{H}_{\mathrm{L}}(\boldsymbol{\theta})$ and $\boldsymbol{H}_{\mathrm{R}}(\boldsymbol{\theta})$ as performed by his HRTF, an estimate of the spectrum of the sound source $\hat{\boldsymbol{S}}$, as well as information on the behavior of the respective errors $\boldsymbol{\delta}_{\mathrm{L}}$, $\boldsymbol{\delta}_{\mathrm{R}}$ and $\boldsymbol{\delta}_{S}$. In the following, we will discuss these *a priori* assumptions and propose quantitative models for them. We acknowledge that these models are simplifications leaving out many detailed observations from spatial hearing research. Nevertheless, as we will show further on, this simplified model will allow us to reproduce and gain insight into essential properties of human sound localization as measured experimentally.

ITD noise model In order to extract sound source direction information from the ITD cues, the listener needs to have a noise model for the ITD measurement, δ_{itd}, i.e., make (*a priori*) assumptions about the distribution of the error $p(\delta_{\mathrm{itd}})$. This error is related to the precision of the hearing apparatus in measuring ITD, which also depends on the frequency content and the intensity of the incoming sound (Hafter and De maio 1975; Kuhn 1977). Indeed, based on a physiological model of the auditory-nerve encoding and a widely accepted cross-correlation model proposed to extract ITD estimates from these neural patterns, it is to be expected that ITD estimates will be affected by the strengths of the acoustic signals received at both ears (Stern and Colburn 1978). Such dependency can also be observed from behavioral experiments (Domnitz 1973; Mossop and Culling 1998), but it is not very pronounced when the stimuli are clearly audible and the ITD–ILD cues match. This same cross-correlation model, through its tonotopically organized multichannel implementation, also argues for the frequency-dependent nature of ITD estimation. Various mechanisms have been proposed to integrate activity across frequency channels and turn the multiple, one for each frequency channel, ITD estimates into a single perceptual ITD cue (Stern et al. 1988; Shackleton et al. 1992; Hancock and Delgutte 2004).

We chose to ignore these intensity and frequency dependencies of ITD cues and to estimate the error distribution based on a single just-noticeable difference (JND) measure extracted from ITD discrimination experiments with clearly audible broadband noise stimuli. As these conditions correspond well with the human localization experiments we will be modeling, we assume this to be a valid approximation for our analysis. The JND is defined as the smallest difference Δx between two cue values (x and $x + \Delta x$) that can be correctly detected for a specified percentage of the experiments. The results of ITD experiments using broadband noise stimuli (Mossop and Culling 1998) show that this JND depends on the particular ITD value around which it is measured. See Hancock and Delgutte (2004) for how this result at the behavioral level can be explained from the responses of a neural population model assuming ITD estimation through cross-correlation in different frequency channels followed by activity integration across frequency channels.

In particular, these experiments show that human ITD discrimination performance is approximately a linear function of ITD, but with a nonzero offset at zero ITD,

$$\mathrm{JND}(\mathrm{itd}) = a + b \cdot \mathrm{itd},$$

with $a = 32.5\ \mu\mathrm{s}$ and $b = 0.095$. With regard to the noise on the ITD, this means that δ_{itd} cannot be described by a single normal distribution, as the noise variance σ^2_{itd} is not constant. Therefore, to simplify the analysis, we first transform the ITD scale expressed in time units into a new scale expressed in JND units,

$$\mathrm{itd} \rightarrow \int_0^{\mathrm{itd}} \frac{1}{\mathrm{JND}(x)}\mathrm{d}x = \frac{1}{b}[\log(a + b \cdot \mathrm{itd}) - \log(a)],$$

where $\mathrm{d}x/\mathrm{JND}(x)$ indicates the number of JND's in the interval $\mathrm{d}x$ and we integrate over x to arrive at the total number of JND's for a given itd. The error on this transformed ITD value can now be modeled as a normal distribution $p(\delta_{\mathrm{itd}}) \sim N(0, \sigma^2_{\mathrm{itd}})$, with a fixed variance, that can be related to the JND obtained in the experiment: $\sigma_{\mathrm{itd}} = 0.569$ JND, see Supplementary Information.

Directional filter information In order to extract the spatial information encoded in the binaural spectrum, the listener should have information on the directional filtering performed by the head and ears, $\boldsymbol{H}_{\mathrm{L}}(\theta)$ and $\boldsymbol{H}_{\mathrm{R}}(\theta)$. We will assume that learning provides humans with a very accurate model of their HRTF. If the listener is suddenly confronted with a different HRTF, e.g., by use of artificial pinnae or by use of a different HRTF in a virtual sound localization experiment, localization performance is degraded until the internal model is adapted to include the new HRTF (Hofman et al. 1998; Van Wanrooij and Opstal 2005).

Spectral noise model The precision of the measurement of the spectrum is limited by the resolution of the hearing apparatus along both the frequency axis and the intensity axis. In the following, we include this in the model (1) by choosing discrete independent frequency channels and (2) by assessing the internal noise distribution in each of the frequency channels. This is done by modeling basic properties of the hearing apparatus, as determined through psychoacoustic experiments in earlier studies.

The spectral representation along the frequency axis is determined by the neural transduction of the movement of the basilar membrane. As a consequence, the spectrum $\boldsymbol{X}$ can be represented (without loss of information) as a vector representing the responses of a limited number of frequency channels, which are considered independent. To choose these so-called frequency channels, we consider frequencies separated by the equivalent rectangular bandwidth (ERB) of the auditory filter, as these frequencies channels are approximately independent. According to Moore and Glasberg (Moore and Glasberg 1983), the ERB (in Hz) is a function of the center frequency f (in kHz) according to

$$\mathrm{ERB}(f) = 6.23 f^2 + 93.39 f + 28.52.$$

We choose the frequency channels such that their bandwidths correspond to the ERB and the corresponding bandpass filters are adjacent but do not overlap. The first frequency channel is at 300 Hz, and has bandwidth $\mathrm{ERB}(f = 0.3\,\mathrm{kHz}) = 57.3\,\mathrm{Hz}$. The next frequency channel is at $f = 357.3\,\mathrm{Hz}$, and so on. Hence, to cover the frequency range [300 Hz–15 kHz], we arrive at 30 frequency channels with non-overlapping rectangular bandwidths.

For the resolution along the logmagnitude axis, we first assume that the intensity of the incoming sound is above the hearing threshold, i.e., in each of the frequency channels the intensity is above the noise level. The measurement of the logmagnitude is independent in each of the frequency channels, and the precision in each of the channels is determined by $\boldsymbol{\delta}_{\mathrm{L}}$ and $\boldsymbol{\delta}_{\mathrm{R}}$, i.e., by the error on the measurement of the left and right spectrum. As was the case for the ITD error, through experience, the listener assesses the precision of his hearing apparatus and hence learns the error distributions $\boldsymbol{\delta}_{\mathrm{L}}$ and $\boldsymbol{\delta}_{\mathrm{R}}$. We model the error on the logmagnitude spectrum due to internal noise as a multivariate normal distribution, given by

$$p(\boldsymbol{\delta}_{\mathrm{L}}) \sim N(0, \boldsymbol{\Sigma}_{\mathrm{L}})$$
$$p(\boldsymbol{\delta}_{\mathrm{R}}) \sim N(0, \boldsymbol{\Sigma}_{\mathrm{R}}),$$

with $\boldsymbol{\Sigma}_{\mathrm{L}} = \boldsymbol{\Sigma}_{\mathrm{R}}$ the monaural covariance matrices. Because of the choice of non-overlapping frequency channels, we assume each of the frequency channels to operate independent of the others, i.e., the covariance matrices are diagonal. Moreover, if we assume that the error distribution is identical in each of the channels, i.e., a normal distribution with variance σ_I^2, then the covariance matrices read $\boldsymbol{\Sigma}_{\mathrm{L}} = \boldsymbol{\Sigma}_{\mathrm{R}} = \sigma_I^2 \cdot \boldsymbol{I}$, where is the unity matrix with order $n_f \times n_f$..

In order to set the parameter of this error distribution σ_I, we again turn to psychoacoustic experiments that tested for intensity discrimination of broad band noise (Hartmann and Constan 2002), see Supplementary information. For our model, using the frequency channels as described above, this results in $\sigma_I = 3.5\,\mathrm{dB}$. Note that using the concept of ERB to determine the 'independent' frequency channels is of course an approximation, not so much for the distribution of frequency channels, but for the exact number of channels. But because the calculated variance in each of the frequency channels σ_I depends on the assumed number of channels, the overall results will not be very sensitive to the exact number of channels. Indeed, if the number of independent channels is actually larger, then the spectral representation can carry more information, but this would be compensated by a higher σ_I which would in turn reduce the information.

Source spectrum information The listener can extract more spatial information from the binaural spectrum, if he has knowledge about the spectrum of the sound source. Based on the incoming signal and *a priori* information, he has to estimate the sound spectrum $\hat{\boldsymbol{S}}$ *and* has to have an idea of how accurate this estimate is, i.e., he should have an idea of the distribution of $\boldsymbol{\delta}_S$. In this paper, we will not try to answer the question how the listener arrives at such an estimate nor assesses the precision of that estimate. We just assume that the spectrum of sound sources can be learned from the ensemble of sounds the listener is regularly confronted with. In that respect, it can be argued that in an evolutionary context, the auditory system has evolved to optimize the localization of transients as these are survival-dependent environmental sounds made by predators or prey. As a transient spectrum can be well approximated by a flat spectrum, it provides a very consistent source spectrum for the auditory system to use. If the listener is very familiar with a certain sound source, it is expected that the error on the source spectrum $\boldsymbol{\delta}_S$ is very small (small variance), which in turn will allow the listener to extract more spatial information from the signal. Indeed, as the noise in the source spectrum is reduced, more weight can be given to the spatial filtering as the cause of the spectral features. On the other hand, if the listener is very unfamiliar with the source, the auditory system should allow the spectral error to be really large (and assume a very large variance), which might hamper sound localization. It is clear that of all the error distributions introduced in the model, the least is known about the assumed distribution of $\boldsymbol{\delta}_S$, as it models to what extent the listener *thinks* he is able to estimate the sound source spectrum, which is hard to quantify. Therefore, we take the freedom to model the presupposed distribution as

$$p(\boldsymbol{\delta}_S) \sim N(0, \boldsymbol{\Sigma}_S).$$

As before, we assume that $\boldsymbol{\Sigma}_S = \sigma_S^2 \cdot \boldsymbol{I}$ with $\boldsymbol{I}$ the $n_f \times n_f$ unity matrix, i.e., all frequency channels are uncorrelated and have identical error variances.

Note that it could be the case that the listener is good at estimating the spectrum of the source, but not the exact intensity of the source. For example, in a particular psychoacoustic experiment, it may be only the relative intensity variations across the binaural spectrum and not so much the absolute intensity of the spectrum that is significant. In order to remove the effect of the unknown sound intensity, one can use a slightly altered covariance matrix given by $\boldsymbol{\Sigma}_S = \sigma_S^2 \cdot \boldsymbol{I} + \sigma^2 \cdot \mathbf{1}$, where σ^2 is the expected variance on the source strength which is a fixed value and $\mathbf{1}$ is the $n_f \times n_f$ unit matrix (all ones). This parameter σ^2 accounts for the covariance of the errors between frequency channels. In the following, we assume that σ is large (5 dB), such that the listener does not assign any meaning to correlated errors and consequently only makes use of relative information in the binaural spectrum, irrespective of the overall intensity of the sound. Simulations for different σ show that the localization performance is rather insensitive to the exact value of σ (results are not shown here).

2.3.3 Bayes' rule revisited

Finally, having introduced the error distributions of $p(\boldsymbol{\delta}_{\text{itd}})$, $p(\boldsymbol{\delta}_{\text{L}})$, $p(\boldsymbol{\delta}_{\text{R}})$ and $p(\boldsymbol{\delta}_S)$ as assumed by the listener, we can write $P(\boldsymbol{X}|\boldsymbol{\theta})$ in the new basis as

$$\boldsymbol{X}_{\boldsymbol{\theta}}[\boldsymbol{\delta}] = \boldsymbol{T}_{\boldsymbol{\theta}} + \boldsymbol{\delta}$$
$$P(\boldsymbol{X}|\boldsymbol{\theta}) = \frac{1}{(2\pi)^{N/2}|\boldsymbol{\Sigma}|^{1/2}} \times \exp\left\{-\frac{1}{2}\left(\boldsymbol{X} - \boldsymbol{T}_{\boldsymbol{\theta}}\right)\boldsymbol{\Sigma}^{-1}\left(\boldsymbol{X} - \boldsymbol{T}_{\boldsymbol{\theta}}\right)\right\}$$

with the angular template $\boldsymbol{T}_{\boldsymbol{\theta}}$

$$\boldsymbol{T}_{\boldsymbol{\theta}} = [\text{itd},(\boldsymbol{H}_{\text{L}} - \boldsymbol{H}_{\text{R}}), (\hat{\boldsymbol{S}} + (\boldsymbol{H}_{\text{L}} + \boldsymbol{H}_{\text{R}})/2)](\theta) \tag{3}$$

and the covariance matrix

$$\boldsymbol{\Sigma} = \begin{bmatrix} \sigma_{\text{itd}}^2 & 0 & 0 \\ 0 & \boldsymbol{\Sigma}_- & 0 \\ 0 & 0 & \boldsymbol{\Sigma}_+ \end{bmatrix}.$$

with $\boldsymbol{\Sigma}_- = 2\sigma_I^2 \cdot \boldsymbol{I}$ and $\boldsymbol{\Sigma}_+ = (\sigma_I^2/2 + \sigma_S^2) \cdot \boldsymbol{I} + \sigma^2 \cdot \mathbf{1}$, which summarizes the *a priori* information on which the localization will be based.

Using Bayes' rule, the posterior probability of source direction $\boldsymbol{\varphi}$ given the measurement $\boldsymbol{X}_{\boldsymbol{\theta}}[\boldsymbol{\delta}]$ originating from a source at direction $\boldsymbol{\theta}$ is given by

$$\begin{aligned} P(\boldsymbol{\varphi}|\boldsymbol{X}_{\boldsymbol{\theta}}[\boldsymbol{\delta}]) &= \frac{1}{P(\boldsymbol{X}_{\boldsymbol{\theta}}[\boldsymbol{\delta}])} P(\boldsymbol{X}_{\boldsymbol{\theta}}[\boldsymbol{\delta}]|\boldsymbol{\varphi}) \cdot P(\boldsymbol{\varphi}) \\ &= C(\boldsymbol{X}_{\boldsymbol{\theta}}[\boldsymbol{\delta}]) \cdot \exp\left\{-\frac{1}{2}\left(\boldsymbol{X}_{\boldsymbol{\theta}}[\boldsymbol{\delta}] - \boldsymbol{T}_{\boldsymbol{\varphi}}\right)^T \right. \\ &\quad \left. \times \boldsymbol{\Sigma}^{-1}\left(\boldsymbol{X}_{\boldsymbol{\theta}}[\boldsymbol{\delta}] - \boldsymbol{T}_{\boldsymbol{\varphi}}\right)\right\} \end{aligned}$$

with the prior probability on the source direction $P(\boldsymbol{\varphi})$ uniformly distributed, and $C(\boldsymbol{X}_{\boldsymbol{\theta}}[\boldsymbol{\delta}])$ a normalization constant such that $\int P(\boldsymbol{\varphi}|\boldsymbol{X}_{\boldsymbol{\theta}}[\boldsymbol{\delta}])\mathrm{d}\boldsymbol{\varphi} = 1$. In order to calculate $P(\boldsymbol{\varphi}|\boldsymbol{X}_{\boldsymbol{\theta}}[\boldsymbol{\delta}])$, we discretize space, to arrive at a limited number of angular templates $\boldsymbol{T}_{\boldsymbol{\varphi}}$. We sample the source directions $\boldsymbol{\varphi}$ uniformly over the total hemisphere, so that every $\boldsymbol{\varphi}_i$ corresponds to a similar solid angle $4\pi/n$, with $n = 2000$ the number of possible different directions (corresponding to $\Delta\boldsymbol{\varphi} = 4.5°$).

Inspection of $P(\boldsymbol{\varphi}|\boldsymbol{X}_{\boldsymbol{\theta}}[\boldsymbol{\delta}])$ itself will turn out to be useful. It allows one to study what source directions run the risk of being confused with what other directions, i.e., which spatial regions are ambiguous. Note that this estimate is based on a single source presentation.

2.3.4 Mean spherical error

Given the above posterior probability, the observer still has to use a decision criterion to decide which direction $\boldsymbol{\varphi}$ is the best guess. The maximum a posteriori estimator is the Bayes decision rule that is optimal in the sense of minimizing the error rate (Duda et al. 2001). Hence, given the acoustic information $\boldsymbol{X}$, its best guess is the direction $\widehat{\boldsymbol{\theta}}(\boldsymbol{X})$, for which

$$P(\widehat{\boldsymbol{\theta}}|\boldsymbol{X}) = \max[P(\boldsymbol{\varphi}|\boldsymbol{X})].$$

This estimate is again based on one single measurement, with one single noise realization $\boldsymbol{\delta}$. In order to evaluate the localization performance, it is necessary to sample the ensemble of noise realizations and perform multiple localization trials, all with different $\boldsymbol{\theta}$ and $\boldsymbol{\delta}$. One of the quantities which can be used to assess localization performance is then given by the mean spherical error

$$\epsilon(\boldsymbol{\theta}) = \int \arccos\left\{\widehat{\boldsymbol{\theta}}(\boldsymbol{X}_{\boldsymbol{\theta}}[\boldsymbol{\delta}]) \cdot \boldsymbol{\theta}\right\} p(\boldsymbol{\delta})\mathrm{d}\boldsymbol{\delta}$$

The mean spherical error $\epsilon(\boldsymbol{\theta})$ is a measure for how accurately one can localize a sound originating from direction $\boldsymbol{\theta}$. This value is expressed in degrees, and the larger this value, the more the estimate will deviate on average from the true source direction. Hence, this parameter quantifies the accuracy of the spatial localization.

As the integral cannot be calculated directly, we use a Monte Carlo approximation

$$\epsilon(\boldsymbol{\theta}) \simeq \frac{1}{M}\sum_{i=1}^{M} \arccos\left\{\widehat{\boldsymbol{\theta}}(\boldsymbol{X}_{\boldsymbol{\theta}}[\boldsymbol{\delta}_i]) \cdot \boldsymbol{\theta}\right\},$$

drawing a large number ($M = 500$) of random samples $\boldsymbol{X}_{\boldsymbol{\theta}}[\boldsymbol{\delta}_i]$ from the measurement distribution for a source at direction $\boldsymbol{\theta}$. In the numerical experiments described below, we use for the generation of the measurement noise $\boldsymbol{\delta}$ the probability density function

$$p(\boldsymbol{\delta}) = C \cdot \exp\left\{-\frac{1}{2}\boldsymbol{\delta}^T \boldsymbol{\Sigma}^{*-1}\boldsymbol{\delta}\right\}$$

with $\boldsymbol{\Sigma}^*$ the covariance matrix of the measurement noise. Note that this covariance matrix is likely to be different from the belief covariance matrix $\boldsymbol{\Sigma}$, which models the *a priori* assumptions of the listener. Ideally, the latter is identical to the ‘real’ covariance matrix. Through learning, the listener can become familiar with the ensemble of sound sources and so $\boldsymbol{\Sigma}$ can converge to $\boldsymbol{\Sigma}^*$. How the listener updates both his templates as well as his belief covariance matrix is outside the scope of the current paper, but there is ample experimental evidence that both elements of the a priori model are subject to learning (Hofman et al. 1998; Dahmen et al. 2010). In the following, we study localization in the context of experiments, where in the different trials the stimulus sound spectra are more or less similar. Therefore, we assume that the belief covariance matrix is identical to the ‘real’ measurement covariance matrix.

We have also calculated the local population response biases $\boldsymbol{b}(\boldsymbol{\theta})$ for different source positions, as was done in Best et al. (2011). This is calculated as

$$\boldsymbol{b}(\boldsymbol{\theta}) = \frac{\sum_{i=1}^{M} \widehat{\boldsymbol{\theta}}(\boldsymbol{X}_{\boldsymbol{\theta}}[\boldsymbol{\delta}_i])}{\left\|\sum_{i=1}^{M} \widehat{\boldsymbol{\theta}}(\boldsymbol{X}_{\boldsymbol{\theta}}[\boldsymbol{\delta}_i])\right\|} - \boldsymbol{\theta},$$

where i again runs over the $M = 500$ random samples.

3 Results

3.1 Mapping ambiguity regions based on posterior probability distributions

In Fig. 2, we have plotted the results of two simulated localization trials (subject 1), when a broadband stimulus ([300 Hz–15 kHz]) is applied. All spherical mappings are done using the Lambert equal area projection. The ITD value is calculated assuming a distance between the ears of 0.15 m. The sound source is considered to be in the far field (>1 m). We have assumed $\sigma_I = 3.5$ dB and for the unknown variations on the source spectrum, we have taken $\sigma_S = \sigma_I = 3.5$ dB.

In the example shown in Fig. 2a, the source is located in the upper frontal hemisphere at $\boldsymbol{\theta} = (-34°, 45°)$, marked with ‘+’ (see Fig. 1 for definition coordinate system).

One notices that the posterior angular distribution $P(\boldsymbol{\varphi}|\boldsymbol{X}_{\boldsymbol{\theta}}[\boldsymbol{\delta}])$ is nonzero for multiple directions and conse-

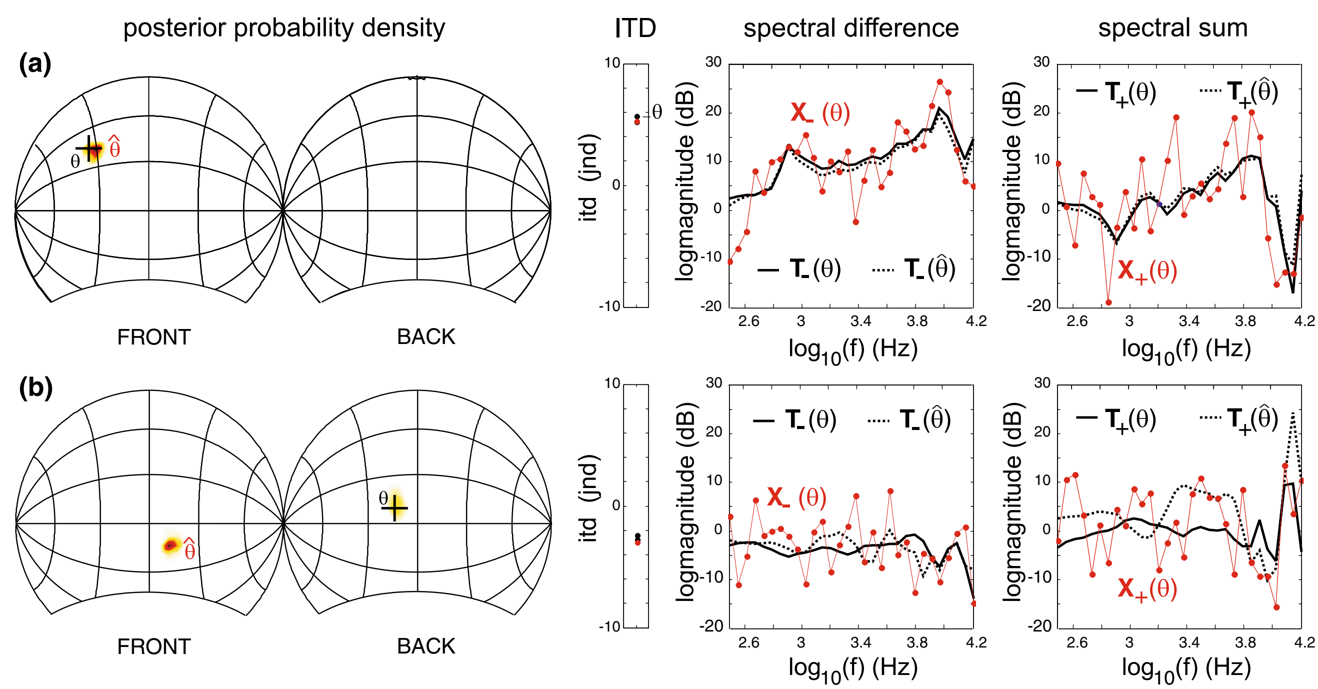

Fig. 2 The results for two simulated trials of a localization experiment for subject 1 (see also Fig. 3a), for a target location in front $\boldsymbol{\theta} = (-34°, 45°)$ (**a**) and in back $\boldsymbol{\theta} = (14°, 171°)$ (**b**) are shown. The figures on the left show the simulated posterior angular probabilities and the true target location (+); the estimated direction corresponds to the direction with maximal posterior probability, respectively $\widehat{\boldsymbol{\theta}} = (-30°, 43°)$ and $(12°, -14°)$. The *figures on the right* show the templates corresponding to the original direction $\boldsymbol{\theta}$ (*solid black line*) and estimated direction $\widehat{\boldsymbol{\theta}}$ (*dotted black line*). The simulated auditory input is shown as the *solid red line*. Note that both the measurement and the template ITD values are close to each other, such that they are plotted on top of each other

quently ambiguities arise. For this particular trial, these ambiguities are close to the source direction, and the maximum of $P(\boldsymbol{\varphi}|\boldsymbol{X}_{\boldsymbol{\theta}}[\boldsymbol{\delta}])$ is attained at $\boldsymbol{\varphi}$ =(-30°,43°) and consequently the source is classified as being located in this direction $\widehat{\boldsymbol{\theta}}$.

The classification is done by comparison of the auditory input with the set of angular templates, which consist of the ITD, the interaural spectral difference ($\boldsymbol{T}_-$) and the interaural spectral sum ($\boldsymbol{T}_+$), see Eq. 3. On the right hand side of the figure, we have plotted the (simulated) auditory input $\boldsymbol{X}_{\boldsymbol{\theta}}[\boldsymbol{\delta}] = \boldsymbol{T}_{\boldsymbol{\theta}} + \boldsymbol{\delta}$ (red curve), together with the templates corresponding with the true $\boldsymbol{T}_{\boldsymbol{\theta}}$ and the estimated direction $\boldsymbol{T}_{\widehat{\boldsymbol{\theta}}}$ (black solid and dotted). Apparently, for this particular noise realization, the auditory input is more likely to correspond to a template of a nearby direction, a misclassification which is likely to arise because templates of nearby directions are similar to each other.

But the localization error can also be much larger. Indeed, in Fig. 2b, the data are shown for a localization trial, where the sound source is positioned in the (upper) rear hemisphere at (14°,171°). In this case, the posterior angular distribution is clearly much less localized, as it extends to the frontal hemisphere, where it even attains its maximum at $\widehat{\boldsymbol{\theta}}$ =(12°,-14°), resulting in a front-back and up-down reversal. The auditory input is better 'explained' by the template corresponding with this distant direction, than with the original one, and consequently, the misclassification is not limited to nearby directions, resulting in much larger localization errors. Note that the ITDs of the original and estimated directions are very similar and so is the measured ITD, which explains why the lateral position is fairly well estimated (only 2° difference). Also, note that the difference between the $\boldsymbol{T}_+$ templates is larger than between the $\boldsymbol{T}_-$ templates, but also that the variance of $\boldsymbol{\delta}_+$ is larger compared to $\boldsymbol{\delta}_-$, a result of the projection to the new ($\boldsymbol{T}_-$, $\boldsymbol{T}_+$) basis, as discussed earlier. Hence, it is mainly the interaural spectral difference $\boldsymbol{T}_-$ that carries most of the spatial information and explains the possible confusion between the two source positions along the polar axis.

In general, the ambiguities are often situated in the vicinity of the source position, extending mostly along the polar axis; the source's lateral position is less ambiguous. As a consequence, the posterior probability distribution can deviate a lot from the correct source direction, resulting mostly in an error along the polar angle, possibly introducing front-back ambiguities.

3.2 Mean spherical error and error bias

In Fig. 3, we have plotted the spherical error, when averaged over a large number of trials, i.e., $M = 500$ noise realizations. The results are plotted for simulations using the HRTFs of 3 different subjects. One notices that the localization performance is different for the 3 subjects. Some are good at localizing sources at positions where others fail. Others, e.g., subject 3, perform worse overall as can be seen from comparing (different scales) individual with average localization performance (see Fig. 4a). There are some similarities, though, and they can be inferred from Fig. 4a, where the localization performance is averaged over 100 subjects. Comparison with the results in Best et al. (2011), also shown in Fig. 4b, shows that the simulation results are in good agreement with the experimental data from the meta-analysis, at least qualitatively. As concluded by the authors of the experimental study, we also find that: 'The localization is worst, by a wide margin, for sounds above and slightly behind the listener.' Furthermore, both studies indicate that optimal performance extends over polar angles from -45° to +30° in the frontal hemisphere. On the other hand, our simulation results indicate that for larger lateral angles, localization accuracy is best for sounds at the back of the listener and slightly worse in front, while in the experiments the opposite is true. The overall magnitude of the errors is also smaller in simulation, compared to the experiments (by a factor 3); note that the performance-color scale is different in Fig. 4a, b. Figure 4 also shows arrows indicating the direction and the size (scaled) of the local population response biases $\boldsymbol{b}(\boldsymbol{\theta})$ for different source positions, as was done in Best et al. (2011).

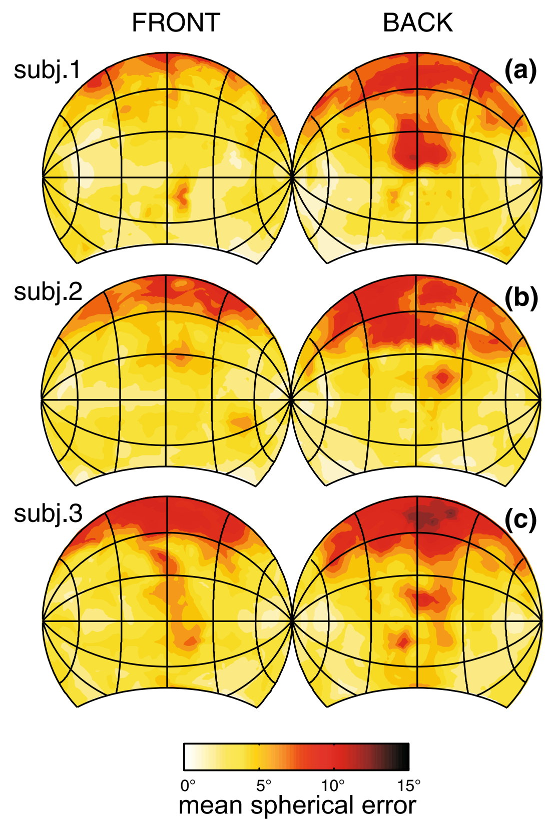

Fig. 3 **a**–**c** The simulated mean spherical error as function of the source position for 3 different subjects. The average was taken over 500 localization trials for each source position

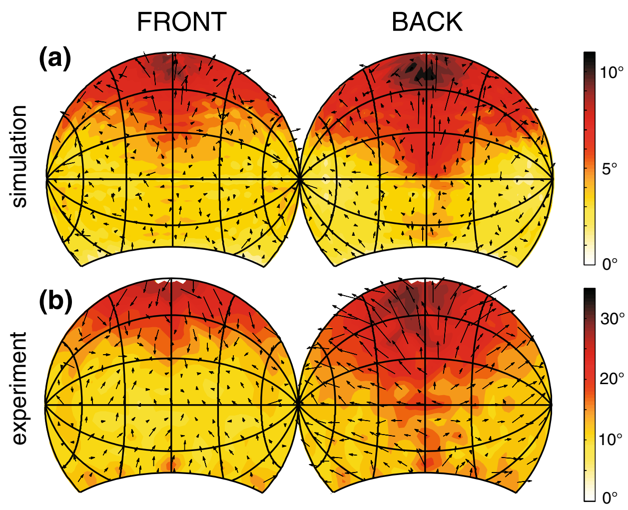

Fig. 4 The mean localization performance (**a**) when the simulation results are averaged over 100 subjects. Superimposed arrows indicate the size and direction of local population response biases for different source positions. **b** The experimentally measured mean localization performance, as reported in Best et al. (2011)

As was the case in the experimental study, we find that in the rear hemisphere, responses are biased forward and toward the left and right poles. For the frontal hemisphere, however, we cannot replicate the experimental finding that the responses are biased toward locations above the horizontal plane. According to our simulation results, most of the bias vectors in this area are small and they point to the rear hemisphere.

The latter discrepancy might be due to the prior distribution, which was not uniform in the experiment (see Best et al. 2011), i.e., not all directions were equally represented. The subjects may have 'learned' this distribution and based their angular estimates on this distribution (see Eq. 2). The bias might also be of strategic nature: It might be a beneficial strategy for the localizer to bias its attention to the lower frontal hemisphere. The observation that the angular error is smaller in simulation compared to that in experiment could be due to the fact that the model makes optimal use of the available information, and consequently, the simulated performance can be interpreted as an upper limit to the actual localization performance; humans are not necessarily ideal observers. But there is also another source of error, which is inherent to the nature of the localization experiments. Indeed, after having localized the sound source in a localization trial, the subject has to communicate his best guess, e.g., by pointing his head, which will introduce an additional localization error. In the meta-analysis, different groups used different reporting procedures, see Best et al. (2011), and consequently, the size of this reporting error cannot be easily assessed.

3.3 Sensitivity analysis of model parameters

The differences between the localization performance in simulation and experiment may also be the result of the specific values chosen for the model parameters. These parameters (except for σ_S) are estimates deduced from psychoacoustical experiments that were conducted under different experimental conditions, possibly resulting in different *a priori* assumptions by the test subjects. Therefore, to find out what effect the uncertainty about the individual model parameters has on the localization performance, we include a sensitivity analysis.

Varying ITD precision: The data are again pooled over 100 subjects. Figure 5a shows the mean spherical error for the model parameters derived above, Fig. 5b for σ_{itd} twice as large (both in the measurement noise and the belief covariance matrix). This changes the performance only very little. Hence, we conclude that for broadband sounds, the localization performance is not too sensitive to the precision of the ITD measurement, at least if σ_S and σ_I are limited, so that the spectra can encode for the lateral angle.

Varying spectral precision: In Fig. 5c, we have reduced the precision of the hearing apparatus for measuring the intensity in each of the frequency channels (σ_I). Doubling σ_I has a large effect and deteriorates performance everywhere. The overall performance degrades, and plenty of front-back localization errors arise, leading to large mean spherical errors. Note that we limited the scale to 20° so that we can compare the results quantitatively with the control situation Fig. 5a.

Varying a priori knowledge on the source spectrum: Of the three model parameters, the parameter σ_S quantifying the variation of the source spectrum is the only one of which we have no experimentally derived estimate. In our simulations so far we have assumed $\sigma_S = \sigma_I$. Doubling σ_S has a clear effect on the localization performance, as can be seen in Fig. 5d. An increase in σ_S actually means that the listener has less knowledge about the source spectrum, which hampers localization especially around the midsagittal plane. Indeed, close to the midsagittal plane (0° lateral angle), the listener can extract less information regarding the polar angle from the ISD ($\boldsymbol{X}_-$) because of symmetry. Therefore, one has to rely more and more on spatial cues encoded by $\boldsymbol{X}_+$. But in order to extract spatial information from $\boldsymbol{X}_+$, one has to have knowledge about the spectrum of the source. However, increasing values of σ_S make that knowledge more uncertain. Therefore, increasing σ_S hampers sound localization, especially for small lateral angles.

Bandwidth of the source spectrum: In the analyses above, the bandwidth of the source signal was assumed to be fixed, rang-

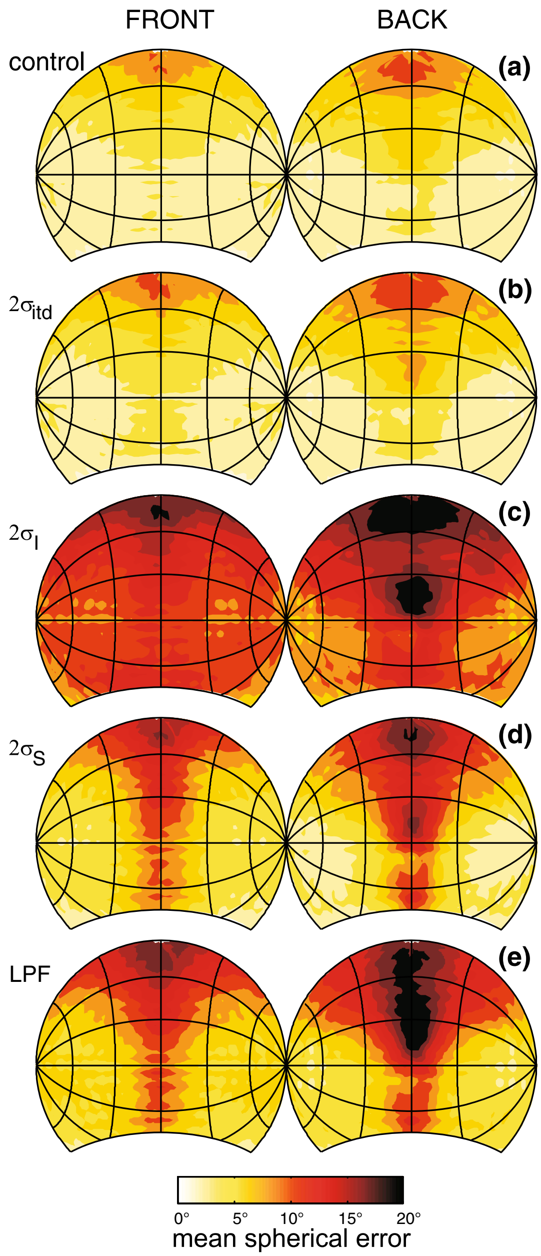

Fig. 5 Sensitivity analysis of the Bayesian localization model. The simulated mean spherical error is shown as function of the source position, when each of the model parameters is varied separately. The average was taken over 500 localization trials for each source position, and 100 subjects were pooled. **a** The model with input parameters as described in the main text. The standard deviation is doubled, respectively, for **b** the noise on the ITD, **c** the internal noise and the variation on **d** the source spectrum. In **e**, the bandwidth of the source is reduced to [300 Hz–8 kHz]

ing from 300 Hz up to 15 kHz. A reduced bandwidth of the perceived sound (due to a reduced bandwidth of the source or due to limitations of the hearing apparatus of the listener, e.g., the loss of sensitivity for the higher frequencies as a result of aging) also affects the localization performance. This can be seen in Fig. 5e, where the localization performance is shown if only frequencies up to 8 kHz are taken into account. This reduces the localization performance significantly over the full sphere, in particular for small lateral angles.

System noise: In the analysis above, we have assumed that the intensity in each of the frequency channels is well above the (system) noise level for each ear, irrespective of the direction the sound is originating from. Hence, the localization errors shown so far, e.g., in the space above the listener, are *not* caused by a limited sensitivity for sounds coming from a particular direction. Similarities between the direction templates corresponding to different source positions explain all the errors. In reality, though, it can happen that the source stimulus is not sufficiently intense for all frequency components to stay above the hearing threshold after directional filtering by the head and ear. Clearly, this loss of spectral information would further degrade the localization performance. In order to test this interaction of stimulus intensity and hearing threshold, we assume a relatively flat stimulus spectrum and vary the signal to noise ratio, see Supplementary Information.

From the result shown in Fig. 6, it is clear that the localization error increases as the intensity of the stimulus decreases (decreasing SNR); the control situation corresponds to SNR = 75 dB. Localization performance decreases faster closer to the midsagittal plane, and the frontal hemisphere outperforms the rear hemisphere, when the SNR decreases. Hence, a limited sound intensity may explain why Best et al. (2011) found that the localization is better in the frontal compared to the rear hemisphere.

4 Discussion

In this paper, we have described an ideal-observer model based on a Bayesian analysis of the localization cues contained in binaurally received signals. The input to this model consists of a binaural HRTF complemented with two parameters that were derived from psychoacoustic experiments measuring the precision of ITD and ILD measurements and one phenomenological parameter that represents the uncertainty about the source spectrum in the listener's mind. Using this model, we were able to explain the sound localization performance in humans, at least qualitatively. A sensitivity analysis of the model parameters shows that various realistic modifications of the different model parameters may account for the observed quantitative discrepancy with the experimental results. Reporting errors may also explain (part of) the discrepancy. Hence, having shown its ability to explain human sound localization in an experimental setup, the new model opens the way for exploring other aspects of sound localization, in other contexts, which we discuss below.

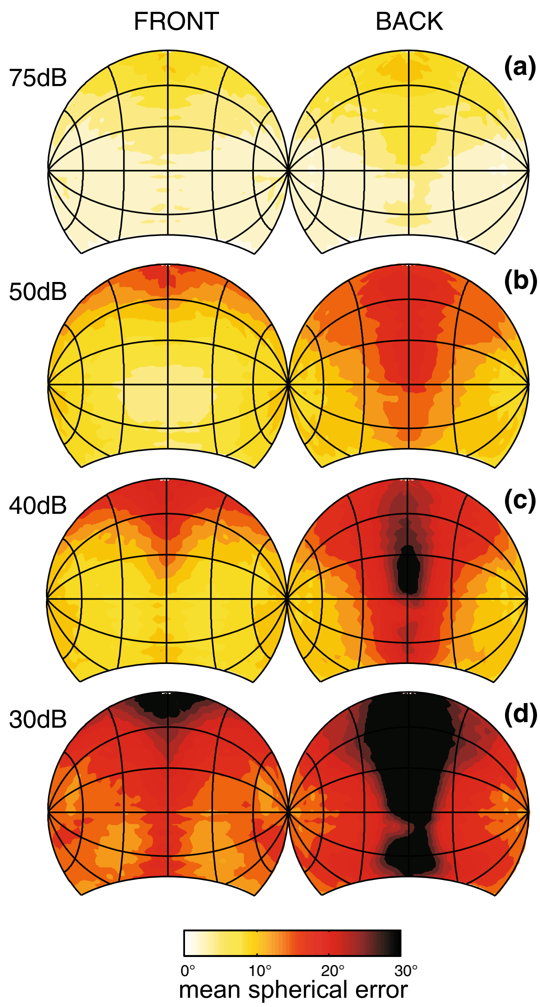

Fig. 6 The mean spherical error for different values of the SNR corresponding to different sound source intensities, see text. SNR = 75 dB corresponds to the control situation, see Fig. 5a, as the magnitude in all frequency channels is above the system noise level

First, the approach described here allows one to study how different aspects of the auditory information (ITD, ILD, ISD, monaural spectral information) interact to result in optimal sound localization. In particular, it allows one to investigate how different spectral distortions, e.g., through processing by a hearing device (cochlear implant) or loss of sensitivity at higher frequencies, etc., could impede correct sound localization. In the case of prosthetic devices and/or virtual 3D sound generation, the model can show how maximal localization capability can be maintained despite encoding limitations by preserving the most informative localization cues contained in the acoustic signals.

Next, the approach could help to shed light on the optimality criterion (and the related loss function) that is used by human observers in a particular context. The proposed model reveals the best localization performance that can be attained by a human subject, i.e., given the constraints of the normal auditory system, assuming a particular optimality criterion. In doing so, the model separates the information carried by the acoustic cues from the strategies and the *a priori* information employed by subjects. The optimality criterion used in our simulations is minimal error rate. In terms of Bayesian decision theory, this corresponds with minimizing the expected loss, with the loss function given by a loss of zero assigned to the correct position and a loss of one assigned to all other positions. Hence, this loss function does not take into account how far off the estimated position is. Other, possibly more appropriate, loss functions can be envisaged but ideally the loss function should be based on the survival value it provides to the decision maker. For example, it can be debated whether a useful loss function penalizes an error in localizing a source behind the subject more than one in front. Some of the qualitative differences between the experimental results and the predictions by the model can perhaps be explained by the experimental subjects using a different loss function. The most important advantage of the proposed model is that it allows considerations at this strategic level to be combined with considerations at the sensory information level into a prediction of actual localization performance.

Furthermore, because of the explicit modeling of *a priori* information, the Bayesian framework also allows one to study the buildup of *a priori* information about the sound source during localization experiments and its effect on sound localization performance. In the model, this prior knowledge is contained within the psychological (belief) covariance matrix and the expected spectrum of the sound source. Optimal performance is achieved when the *a priori* information corresponds with the real covariance matrix and the real source spectrum, requiring adaptation to the actual stimulus ensemble. Most likely the listener makes use of heuristics, but one would expect that a listener, confronted with a new ensemble of sound stimuli, would start off with a rather general belief covariance matrix, with large uncertainties on the source spectrum. As the ensemble is increasingly sampled over consecutive trials and the listener 'learns' the ensemble of sound stimuli, his belief covariance matrix will converge gradually (or collapse in case the ensemble is well known) to the real covariance matrix. It would be interesting to challenge the constructed belief system, by occasionally applying unexpected sound stimuli which do not match the learned *a priori* assumptions of the listener. The erroneous sound localizations resulting from those trials could then be compared with the predictions from the model.

In a similar way, one can also study the effect of *a priori* information about the position of the source. In the presented model, it is assumed that the listener has no *a priori* information on the position of the sound source resulting in a uniform prior. However, in reality, many sources generate multiple sounds and consequently one localization event can often build on previous localizations. Indeed, assuming the noise contributions in consecutive localization trials of the same sound source independent, the posterior probability distribution derived by the model can act as a more informed, i.e., non-uniform, prior for interpreting the next localization trial. But also in localization experiments where single sound localization is the norm, the listener often has some *a priori* information on the source position, since space is not sampled uniformly, see Best et al. (2011). According to our model, this *a priori* information will also affect the localization performance and should be taken into account when interpreting the experimental results.

Acknowledgments We would like to thank Virginia Best for providing us with Fig. 1 and the data, used in the meta-analysis reported on in Best et al. (2011). We would also like to thank the reviewers for their thorough reading of the manuscript. Their comments and suggestions clearly improved the manuscript. We acknowledge the support of the Australian Research Council Discovery Project funding scheme (ARC DP110102920).

References

Algazi VR, Avendano C, Duda RO (2001) Elevation localisation and head-related transfer function analysis at low frequencies. J Acoust Soc Am 109:1110–1122

Best V, Kalluri S, McLachlan S, Valentine S, Edwards, Carlile S (2010) A comparison of CIC and BTE hearing aids for three-dimensional localisation of speech. Int J Audiol 1–10 (early online)

Best V, Brungart D, Carlile S, Jin C, Macpherson E, Martin R, McAnally K, Sabin A, Simpson B (2011) A meta-analysis of localisation errors made in the anechoic free field. In: Principles and applications of spatial hearing. World Scientific, Singapore, pp 14–23

Blauert J (1997) Spatial hearing: the psychophysics of human sound localisation. MIT Press, Cambridge, MA

Boyd A, Whitmer W, Soraghan J, Akeroyd M (2012) Auditory externalization in hearing-impaired listeners: the effect of pinna cues and number of talkers. J Acoust Soc Am 131:268–274

Brungart DS, Rabinowitz WM (1999) Auditory localization of nearby sources. Head-related transfer functions. J Acoust Soc Am 106:1465–1479

Carlile S, Pralong D (1994) The location-dependent nature of perceptually salient features of the human head-related transfer function. J Acoust Soc Am 95:3445–3459

Colburn HS (1973) Theory of binaural interaction based on auditory-nerve data. I. General strategy and preliminary results on interaural discrimination. J Acoust Soc Am 54(6):1458–1470

Colburn HS, Kulkarni A (2005) Models of sound localisation. In: Popper AN, Fay RR (eds) Sound source localisation. Springer, Berlin

Dabak AG, Johnson DH (1993) Function-based modeling of binaural processing: level and time cues. J Acoust Soc Am 94(5):2604–2616

Dahmen JC, Keating P, Nodal FR, Schulz AL, King AJ (2010) Adaptation to stimulus statistics in the perception and neural representation of auditory space. Neuron 66:937–948

Domnitz R (1973) The interaural time jnd as a simultaneous function of interaural time and interaural amplitude. J Acoust Soc Am 53(6):1549–1552

Duda RO, Hart PE, Stork DG (2001) Pattern classification, 2nd edn. Wiley, New York

Hafter ER, De maio J (1975) Difference thresholds for interaural delay. J Acoust Soc Am 57:181–187

Hancock KE, Delgutte B (2004) A physiologically based model of interaural time difference discrimination. J Neurosci 24:7110–7117

Hartmann WM, Constan ZA (2002) Interaural level differences and the level-meter model. J Acoust Soc Am 112:1037–1045

Hofman PM, Van Opstal AJ (2003) Binaural weighting of pinna cues in human sound localization. Exp Brain Res 148:458–470

Hofman PM, Van Riswick JG, Van Opstal AJ (1998) Relearning sound localisation with new ears. Nat Neurosci 1:417–421

Jin C, Corderoy A, Carlile S, van Schaik A (2004) Contrasting monaural and interaural spectral cues for human sound localisation. J Acoust Soc Am 115:3124–3141

Kuhn GF (1977) Model for the interaural time differences in the azimuth plane. J Acoust Soc Am 62:157–167

Kulkarni A, Colburn HS (1998) Role of spectral detail in sound-source localisation. Nature 396:747–749

Langendijk EH, Bronkhorst AW (2002) Contribution of spectral cues to human sound localisation. J Acoust Soc Am 112:1583–1596

Macpherson EA (1997) A comparison of spectral correlation and local feature-matching models of pinna cue processing. J Acoust Soc Am 101:3104

Macpherson EA, Sabin AT (2007) Binaural weighting of monaural spectral cues for sound localization. J Acoust Soc Am 121(6):3677–3688

Middlebrooks J (1992) Narrow-band sound localisation related to external ear acoustics. J Acoust Soc Am 92:2607–2624

Moore BCJ, Glasberg BR (1983) Suggested formulae for calculating auditory-filter bandwidths and excitation patterns. J Acoust Soc Am 74:750–753

Mossop JE, Culling JF (1998) Lateralization of large interaural delays. J Acoust Soc Am 104:1574–1579

Musicant AD, Butler RA (1984) The influence of pinnae-based spectral cues on sound localisation. J Acoust Soc Am 75:1195–1200

Shackleton TM, Meddis R, Hewit MJ (1992) Across frequency integration in a model of lateralization. J Acoust Soc Am 91(4):2276–2279

Shinn-Cunningham BG, Santarelli S, Kopco N (2000) Tori of confusion: binaural localization cues for sources within reach of a listener. J Acoust Soc Am 107:1627–1636

Stern RM, Colburn HS (1978) Theory of binaural interaction based on auditory-nerve data. IV. A model for subjective lateral position. J Acoust Soc Am 64(1):127–140

Stern RM, Zeiberg AS, Trahiotis C (1988) Lateralization of complex binaural stimuli: a weighted-image model. J Acoust Soc Am 84:156–165

Van Wanrooij MM, Van Opstal AJ (2005) Relearning sound localisation with a new ear. J Neurosci 25:5413–5424

Wightman FL, Kistler DJ (1989) Headphone simulation of free-field listening II: psychophysical validation. J Acoust Soc Am 85:868–878

Woodworth R (1938) Experimental psychology. Holt, New York