Sound source localization in real sound fields based on empirical statistics of interaural parameters[a)]

Johannes Nix[b)] and Volker Hohmann
Medizinische Physik, Carl von Ossietzky Universität Oldenburg, D-26111 Oldenburg, Germany

(Received 12 July 2004; revised 20 October 2005; accepted 24 October 2005)

The role of temporal fluctuations and systematic variations of interaural parameters in localization of sound sources in spatially distributed, nonstationary noise conditions was investigated. For this, Bayesian estimation was applied to interaural parameters calculated with physiologically plausible time and frequency resolution. Probability density functions (PDFs) of the interaural level differences (ILDs) and phase differences (IPDs) were estimated by measuring histograms for a directional sound source perturbed by several types of interfering noise at signal-to-noise ratios (SNRs) between −5 and +30 dB. A moment analysis of the PDFs reveals that the expected values shift and the standard deviations increase considerably with decreasing SNR, and that the PDFs have non-Gaussian shape at medium SNRs. A d' analysis of the PDFs indicates that elevation discrimination is possible even at low SNRs in the median plane by integrating information across frequency. Absolute sound localization was simulated by a Bayesian maximum *a posteriori* (MAP) procedure. The simulation is based on frequency integration of broadly tuned "detectors." Confusion patterns of real and estimated sound source directions are similar to those of human listeners. The results indicate that robust processing strategies are needed to exploit interaural parameters successfully in noise conditions due to their strong temporal fluctuations. [DOI: 10.1121/1.2139619]

PACS number(s): 43.66.Ba, 43.66.Pn, 43.66.Qp, 43.72.Dv [AK] Pages: 463–479

I. INTRODUCTION

The filtering of acoustical signals by the human body, head, and pinna is characterized by the head-related transfer functions (HRTFs) and depends on both direction and distance of the sound source (Blauert, 1983; Shaw, 1997). A set of HRTFs for the left and right ears characterizes the physical differences between signals recorded at the ear canal entrances. These differences are generally quantified by the frequency-dependent interaural parameters, i.e., the interaural level differences (ILDs) and the interaural time differences (ITDs). Interaural parameters measured from different directions exhibit a rich variety of possible features for sound localization (Wightman and Kistler, 1989b) and have therefore been considered in many physiological and psychoacoustical models of binaural processing. Jeffress (1948) proposed a "place theory" of sound localization based on the ITDs, which suggests a physiological mechanism for coincidence detection. Influenced by Jeffress's hypothesis, ITDs have been used in many psychoacoustic experiments and models of binaural detection to characterize interaural timing (Breebaart *et al.*, 1999; Colburn, 1996; Durlach and Colburn, 1978). Responses of neurons to ITDs were also considered in physiological studies of binaural processing (Brugge, 1992; Caird and Klinke, 1987; Joris and Yin, 1996; Clarey *et al.*, 1992; Kuwada and Yin, 1987). Alternatively, interaural phase differences (IPDs), which are the frequency-domain representation of ITDs, have been used to quantify interaural timing cues (Kuwada and Yin, 1983; Malone *et al.*, 2002; Spitzer and Semple, 1991). IPDs were used as well in recent quantitative physiological models (Borisyuk *et al.*, 2002). Whether IPD or ITD representations of interaural timing cues are more physiologically relevant for auditory processing in mammals is still an open question (McAlpine and Grothe, 2003). Regarding the processing of ILDs, there is a wide consensus that a combination of excitatory ipsilateral and inhibitory contralateral interactions takes place (Colburn, 1996). Interaural timing information and ILDs are then combined for sound localization (Brugge, 1992). In the barn owl (tyto alba) it has been shown that a topotopic map of auditory space exists, which performs such a combination (Knudsen, 1982).

Interaural parameters have also been used as basic parameters for sound source localization algorithms (Neti *et al.*, 1992; Albani *et al.*, 1996; Datum *et al.*, 1996; Duda, 1997; Janko *et al.*, 1997; Chung *et al.*, 2000; Liu *et al.*, 2000), "cocktail-party" processors (Bodden, 1993), and binaural directional filtering algorithms (Kollmeier *et al.*, 1993; Kollmeier and Koch, 1994; Wittkop *et al.*, 1997). The aim of such algorithms is to estimate the directions of the sound sources on a short-term basis and use this information for speech enhancement techniques like Wiener filtering (Bodden, 1996). A short-term analysis of interaural parameters is commonly used in these approaches. It is also assumed that the auditory system initially evaluates interaural parameters on a short-term basis for exploiting binaural information.

[a)]Part of this research was presented at the 137th meeting of the Acoustical Society of America [V. Hohmann and J. Nix, "Application of localization models to noise suppression in hearing aids," J. Acoust. Soc. Am.**105**, 1151 (1999)].

[b)]Electronic mail: johannes.nix@uni-oldenburg.de

For localization of signals in noise or of nonstationary signals, the information available for sound localization differs from the information in the HRTFs. In contrast to the stationary and anechoic conditions in which HRTFs are measured, the interaural parameters derived from subsequent windows in a short-term analysis fluctuate due to the statistics of the signal and due to the influence of noise, if present. As a consequence, the information about source location in the short-term values of the interaural parameters is likely to be degraded as compared to the information content of the set of HRTFs itself. These fluctuations have a close relationship with the properties of the HRTFs, e.g., spectral notches of the HRTFs for certain directions can lead to stronger fluctuations at these frequencies. Therefore, fluctuations can be used to retrieve additional information on the sound source direction.

Probability density functions of interaural parameters have been evaluated in several studies on binaural detection (Domnitz and Colburn, 1976; Henning, 1973; Zurek, 1991). However, these experiments do not allow us to draw conclusions about localization performance in noisy environments. Although fluctuations of interaural parameters have been explicitly considered in models of sound localization (Duda, 1997; Wittkop *et al.*, 1997), empirical data on the amount of fluctuations are still missing. The aim of the present study is to characterize fluctuations of interaural parameters in realistic listening conditions and to study their possible implications for sound localization.

The approach chosen here is to simulate the possible effective signal processing involved in binaural sound localization and to use *a priori* knowledge about statistical properties of the interaural parameters to estimate the sound source directions. This *a priori* knowledge is gathered empirically by observation of histograms of a large amount of short-term interaural parameter values. Consequences of the observed statistics for modeling sound source localization are discussed using the framework of detection theory and Bayesian analysis. It was not in the focus of these analyses to construct a detailed model of human sound localization, but to gain knowledge about the relevant properties of real-world signals.

II. METHODS

A. Assumptions

Several general assumptions are made in this study: First, only binaural cues derived from the amplitude ratio and the phase difference of the left and right signals are used for simulating sound source localization.

Second, the sound to be localized is assumed to be speech coming from a constant direction without reverberation. The noise is assumed to be a noise field as found in real environments without preferred directions and to be incoherent between different frequencies. The long-term signal-to-noise ratio is known, however the shape of the short-term magnitude spectra of the participating sound sources is not known.

The global structure of the processing is in line with current auditory peripheral modeling approaches: First, a short-term frequency analysis with resolution similar to the auditory critical bandwidth is performed. Interaural parameters are then calculated for the set of frequency bands as a function of time. Second, binaural information is integrated across the different frequency bands. In contrast to most models mentioned in the Introduction, this integration is defined as a combination of probabilities. Because time domain and frequency domain representation of a signal are equivalent in terms of information content, the IPDs as frequency-domain representation of the interaural timing carry approximately the same information as the narrow-band ITDs. To represent interaural timing, the IPDs are used here in addition to the ILDs. Finally, a temporal integration with a longer time constant is performed. The time constants of the model are a 16-ms window for the analysis of interaural parameters, followed by a moving average with 8-ms time constant, and, subsequent to the computation of probabilities, a 100-ms window for the integration of statistical information, similar to the time constants found by Wagner (1991). The long-term integration consists of a simple moving average; statistical estimation procedures of a much higher complexity are possible and may be required to explain localization of several concurrent sound objects. However, they are disregarded here because this study focuses on localization of a single directional source.

Using short-term analysis of nonstationary signals to evaluate interaural parameters has important consequences on the amount of temporal fluctuations, especially if framing prior to ITD detection is assumed. As an example, Fig. 1 shows schematically the time course of the level at the left

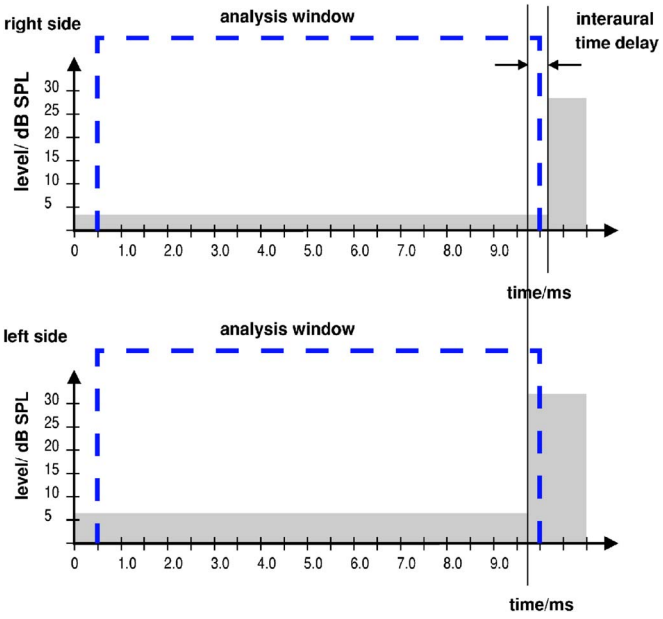

FIG. 1. (Color online) The diagram shows the effect of fluctuations of a single signal with a fixed interaural time delay on the observed ILDs for the framework of a short-term analysis of both ear signals. As long as the signal is constant, an ILD of about 3 dB is observed. In the example, a sudden increase in level occurs close to the end of the analysis window. This increase is caught in the temporal window of the ipsilateral side, but not the one of the contralateral side. Because the ILD and IPD values are weighted according to their amplitude, the higher ILD near the end of the analysis window gains a high weight. A temporal fluctuation of the observed short-term ILDs is the result.

and right ear canal entrance. The source is assumed to be on the left side. Therefore, we usually expect the level at the left side to be higher than the level at the right side, resulting in a stationary level difference of about 3 dB for our example. Now consider what happens if the signal has a sudden increase in level of about 25 dB near the end of the analysis window. At the left side, part of the additional energy enters the short-term analysis window, but not at the right side. Because the increase by 25 dB means a more than tenfold increase in the amplitude, the observed intensity-weighted level difference is much larger than one would expect because of the sound source direction. The same is valid for level-weighted estimates of interaural timing parameters, when analyzed by a windowed short-term analysis.

B. Computation of interaural parameters in the frequency domain

In this work, the interaural parameters are represented in the frequency domain by the interaural transfer function (ITF) (Duda, 1997). It is defined by the HRTFs, $H_r(\alpha,\phi,f)$ and $H_l(\alpha,\phi,f)$, which are functions of azimuth α, elevation ϕ, and frequency f; r and l denote left and right side. As the elevation of a sound source, we define the angle to the horizontal plane; as the azimuth, we define the angle between the projection of the source onto the horizontal plane, and the median plane. The ITF is defined as the quotient of the left and right HRTFs, assuming both have nonzero gains:

$$I(\alpha,\phi,f)=\frac{H_r(\alpha,\phi,f)}{H_l(\alpha,\phi,f)}. \tag{1}$$

Interaural timing can be characterized by the phase of the ITF,

$$\arg I(\alpha,\phi,f), \tag{2}$$

by the *interaural phase delay* t_p,

$$t_p(\alpha,\phi,f)=-\frac{\arg I(\alpha,\phi,f)}{2\pi f}, \tag{3}$$

and by *the interaural group delay*, t_g,

$$t_g(\alpha,\phi,f)=-\frac{1}{2\pi}\frac{d\arg I(\alpha,\phi,f)}{df}. \tag{4}$$

If one nonstationary sound source with spectrum $S(f,t)$ is present, t denoting the time variable, the spectra of the signals which arrive, filtered by the HRTFs, at the left and right ear canals are approximately

$$F_r(\alpha,\phi,f,t)\approx S(f,t)H_r(\alpha,\phi,f), \tag{5}$$

$$F_l(\alpha,\phi,f,t)\approx S(f,t)H_l(\alpha,\phi,f). \tag{6}$$

We do not write this as an identity because of the possible windowing effects for nonstationary sources discussed above.

If we assume $F_l(\alpha,\phi,f,t)$ and $F_r(\alpha,\phi,f,t)$ to be nonzero, the time-dependent quotient transfer function $\hat{I}(\alpha,\phi,f)$ can be computed based on the binaural signals as an approximation to the ITF:

$$\hat{I}(\alpha,\phi,f,t)=\frac{F_r(\alpha,\phi,f,t)}{F_l(\alpha,\phi,f,t)}. \tag{7}$$

Here, the ILD Δ_L and the IPD Δ_θ are defined as amplitude and phase of $\hat{I}(\alpha,\phi,f)$.

Because, with X^* being the complex conjugate of X,

$$\hat{I}(\alpha,\phi,f,t)=\frac{F_r(\alpha,\phi,f,t)F_l(\alpha,\phi,f,t)^*}{|F_l(\alpha,\phi,f,t)|^2}, \tag{8}$$

the IPDs can be computed from the interaural cross-power spectrum without unwrapping. $F_r(\alpha,\phi,f,t)/F_l(\alpha,\phi,f,t)$ is, according to the cross-correlation theorem (a generalized form of the Wiener-Khintchine theorem), identical to the Fourier transform of the short-term interaural cross-correlation function.

C. Description of interaural parameters as random variables

It is assumed that short-term interaural parameters form a stochastic process. It is described by the vector random variable of interaural parameters $\vec{\boldsymbol{\Delta}}$ whose instances $\vec{\Delta}$ consist of the set of ILD and IPD variables in all frequency bands at a certain point of time. Writing $\Delta_{L,b}$ for the ILD of band b, and $\Delta_{\theta,b}$ the IPD of band b, $\vec{\Delta}$ is defined as

$$\vec{\Delta}=(\Delta_{L,1},\Delta_{L,2},\ldots,\Delta_{L,B},\Delta_{\theta,1},\Delta_{\theta,2},\ldots,\Delta_{\theta,B}). \tag{9}$$

Assuming that B frequency bands are analyzed, $\vec{\Delta}$ is $2B$-dimensional. In the following, we disregard the temporal characteristics of $\vec{\boldsymbol{\Delta}}$, e.g., temporal correlations, and focus on the probability density function (PDF) of $\vec{\boldsymbol{\Delta}}$. The PDF of the random variable $\vec{\boldsymbol{\Delta}}$ is determined by the properties of the sound field at both ears. In principle, this PDF could be calculated analytically, requiring knowledge about the anechoic HRTFs, the power spectral density statistics of the sources, room acoustics, and the distribution of incidence directions of the noise sources. In practice, however, an analytical derivation is not feasible because the required statistics are not available.

In this study, the PDF is therefore estimated empirically by measuring the time series of short-term interaural parameters calculated from actual binaural recordings of directional sound sources. The normalized histogram of the series is then regarded as being an estimate of the PDF of the underlying random process. We denote the PDF of $\vec{\boldsymbol{\Delta}}$ given the direction λ as $p(\vec{\Delta}|\lambda)$, where λ is one of N_λ possible discrete directions. As $p(\vec{\Delta}|\lambda)$ is a $2B$-dimensional PDF, and histograms with N_K categories require about $(N_K)^{2B}$ observations, its empirical estimation is not feasible for $B=43$, because the number of observations required would be too high. Therefore, only the histograms of the components of $\vec{\Delta}$ are observed, and it is assumed that the components of $\vec{\boldsymbol{\Delta}}$ can be treated as statistically independent. In this case, the joint PDF is calculated by multiplication of the component PDFs

$$p(\vec{\Delta}|\lambda) = \prod_{b=1}^{2B} p_b(\Delta_b|\lambda), \tag{10}$$

where the component or marginal PDFs $p_b(\Delta_b|\lambda)$ are estimated by the histograms of the respective variables Δ_b.

In addition to approximating the PDF of $\vec{\mathbf{\Delta}}$ by the product of its marginals, stationarity, and ergodicity of the random process are assumed. Because the statistics of spectral power densities of speech cannot be assumed to be Gaussian, and interaural parameters are the result of a nonlinear operation on the spectral power densities, it can not be assumed, in general, that the resulting PDF has a Gaussian shape.

D. Data acquisition

1. Signals

The signals used in this study are an additive superposition of one directional sound source recorded in an anechoic room (the target signal) and spatially distributed noises recorded in real listening environments.

All signals were recorded binaurally in one individual male subject using hearing aid microphones mounted in in-the-ear (ITE) hearing aids (Siemens Cosmea M). The devices were fitted to the subject's ear canals with ear molds, allowing for a reproducible positioning of the microphones. They were positioned at the ear canal entrance, ensuring that most of the binaural information would be recorded; however, the effect of the concha may have been underestimated due to the size of the devices. The usable frequency range of the hearing aid microphone was between 100 Hz and about 8 kHz.

A nonstationary speech signal was generated as the target signal. Continuous speech from four different talkers was recorded at a sampling frequency of 25 kHz through a single microphone and added together to yield a speech signal with almost no pauses but still significant amplitude modulations. It included one female and three male talkers. One male read a German text and the other talkers read English texts. From this recording a segment of 20-s duration was selected as a target in which all talkers were present.

In the next step, directional information was generated by taking binaural recordings of the target signal in an anechoic room. The position of the source was set by means of a free-field apparatus (Otten, 2001). The apparatus consists of two loudspeakers attached to a movable arc of 3.6-m diameter, which can be positioned with stepping motors at all azimuths and at all elevations between −30° and 85°. Positions of the loudspeakers and generation of signals were controlled by a personal computer. The subject sat in the center of the arc on a stool and was instructed to move his head and arms as little as possible. The head was supported by a headrest. The error of direction is estimated to be about 3° in the azimuth and elevation, mostly the result of head movements and to a lesser extent due to position uncertainty. Recordings of the target signal on digital audio tape (DAT) were taken at 430 positions, ranging in azimuths from 0° to 355° in 5° steps at elevations of −15°, 0°, 15°, and 30°. For azimuths in the range of −15° to 15° and 165° to 195°, additional elevations of −10°, −5°, 5°, 10°, and 20° were measured. Three breaks were scheduled during the session in order to change the tapes and to permit the subject to take a rest and move around. All elevations for each azimuth were recorded without a break.

TABLE I. List of the 27 signal conditions chosen for the analysis of interaural parameters. Each condition includes a superposition of a directional target signal from 430 directions and a spatially distributed noise at a specific signal-to-noise ratio. The target signal in silence was included as well.

Target signal	Noise	SNR/dB
speech	silence	⋯
	station concourse	15, 5
	cafeteria	30, 20, 15, 10, 12, 5, 3, 2, 1, 0, −1, −2, −5
	metal workshop 1	15, 5
	metal workshop 2	15
	car interior noise	15, 5
	outdoor market	15, 5
	traffic noise 1	15, 5
	traffic noise 2	15, 5

In addition to the recordings of the target signal, spatially distributed noise signals were recorded in eight different real environments. By using the same subject, the same microphones, and the same recording equipment, it was assured that the anechoic HRTFs and the transfer functions of the equipment were equal in all recordings. The selected noise environments were a densely populated university cafeteria, an outdoor market, inside a moving car, a city street with traffic noise (two different recordings), a train station concourse, and a metal workshop with various machinery operating, such as a milling cutter, grinding wheel, and lathe (two different recordings). The goal was to record situations in which many nonstationary noise sources were impinging on the listener from different directions at the same time and in which no source was dominant for more than about 1 s. The level of the noise was not measured during the recording session but was ensured to be well above the noise floor associated with the recording equipment so that the interaural parameters were determined by the environmental noise rather than the recording noise. Segments of 25-s duration which met these criteria were selected as spatially distributed noise samples.

The DAT recordings of the target and all noise signals were sampled at 25 kHz with 16-bit resolution and stored on a computer. For further processing and analysis of the interaural parameters, 27 different target-noise conditions were selected. Each condition consists of a set of 430 signals covering the different directions of incidence. They were generated by digitally adding a target and a noise signal at various signal-to-noise ratios (SNRs, defined explicitly in the next paragraph) in the range between +30 and −5 dB as well as in silence. The conditions are listed in Table I. Especially for the condition of a speech target mixed with cafeteria noise, a broad range of SNRs were selected to ensure good coverage of this important communication situation.

The SNR was defined as the difference in decibels of the level associated with the target and noise signals. The levels

were calculated from the digitized recordings by averaging the overall rms (root mean square) levels across both ears on a linear scale and transforming this average to decibels. For a specified SNR, the recorded signals were scaled appropriately for each direction, target, and noise type and then added. Using this definition, the SNR was controlled at ear level and did not vary with direction. As the SNR is expected to influence the distribution of interaural parameters, this definition seems more appropriate than defining the SNR at the source level which would have introduced SNR variations with direction due to the influence of the HRTFs. This issue will be discussed later on in this paper.

2. Calculation of interaural parameters and their distributions

Interaural parameters, i.e., ILDs and IPDs, were calculated using a windowed short-term fast Fourier transform (FFT) analysis (Allen and Rabiner, 1977; Allen, 1977) with a subsequent averaging across broader frequency bands.

Time segments with a length of 400 samples were taken from the left and right signals with an overlap of 200 samples. This corresponds to a total window duration of 16 ms and a window time shift of 8 ms, resulting in a frame rate of 125 Hz. The segments were multiplied by a Hann window, padded with zeros for a total length of 512 samples and transformed with a fast Fourier transform (FFT). The short-term FFT spectra of left and right channels are denoted as $F_l(f,k)$ and $F_r(f,k)$, respectively. The indices f and k denote the frequency and time index of the spectrogram, respectively. The FFT bins were grouped to 43 adjacent frequency channels so that, according to the transformation from frequency to the equivalent rectangular bandwidth (ERB) of auditory filters given by Moore (1989), a bandwidth of at least 0.57 ERB was reached. These frequency bands covered the range of center frequencies from 73 Hz to 7.5 kHz, i.e., 3.3 to 32.6 ERB. Because at low frequencies the frequency resolution of the FFT, 48.8 Hz, does not allow for a channel bandwidth of 0.57 ERB, the low-frequency channels have bandwidths of up to 1.37 ERB; the first band which includes more than one FFT bin has a center frequency of 634 Hz, and the average bandwidth is 0.76 ERB.[1]

Let $f_u(b)$ and $f_h(b)$ denote the lowest and highest FFT frequency index belonging to a frequency channel b, respectively. Frequency averaging was then performed by adding up the squared magnitude spectrum and the complex-valued cross spectrum of left and right FFT results across the FFT bins belonging to each channel:

$$F_{rr}(b,k) = \sum_{f=f_u(b)}^{f_h(b)} |F_r(f,k)|^2, \tag{11}$$

$$F_{ll}(b,k) = \sum_{f=f_u(b)}^{f_h(b)} |F_l(f,k)|^2, \tag{12}$$

$$F_{rl}(b,k) = \sum_{f=f_u(b)}^{f_h(b)} F_r(f,k)F_l(f,k)^*. \tag{13}$$

These parameters were filtered by a recursive first-order low-pass filter with the filter coefficient $\gamma = e^{-\Delta T/\tau_a}$, corresponding to a time constant of $\tau_a = 8$ ms, the frame shift being $\Delta T = 8$ ms:

$$\overline{F_{rr}(b,k)} = (1-\gamma)F_{rr}(b,k) + \gamma\overline{F_{rr}(b,k-1)}, \tag{14}$$

$$\overline{F_{ll}(b,k)} = (1-\gamma)F_{ll}(b,k) + \gamma\overline{F_{ll}(b,k-1)}, \tag{15}$$

$$\overline{F_{rl}(b,k)} = (1-\gamma)F_{rl}(b,k) + \gamma\overline{F_{rl}(b,k-1)}. \tag{16}$$

The binaural parameters, i.e., the ILD Δ_L and IPD Δ_θ, were then calculated as follows:

$$\Delta_L(b,k) = 10 \log \left| \frac{\overline{F_{rr}(b,k)}}{\overline{F_{ll}(b,k)}} \right|, \tag{17}$$

$$\Delta_\theta(b,k) = \arg \overline{F_{rl}(b,k)}. \tag{18}$$

It should be noted that Eq. (13) describes an intensity-weighted average of the phase difference across the FFT bins belonging to one channel. The cyclical nature of the IPDs is accounted for by taking the complex-valued average.

The time series of the binaural parameters from each signal were sorted into histograms employing 50 nonzero bins in an adjusted range from up to a −50 to 50 dB level difference and up to a $-\pi$ to $+\pi$ phase difference, respectively. No special effort was made to detect outliers, except speech pauses that were discarded using a threshold criterion.

In each of the 27 different target-noise conditions described earlier, histograms of the ILD and IPD variables in each of the 43 frequency channels for each of the 430 directions were calculated, resulting in a total of 36 980 histograms per condition. The histograms were normalized so that they could be used as an estimate of the marginals of the PDF of the underlying process.

The processing described above was implemented on a VME-Bus-based digital signal processor (DSP) system, made up of five DSPs (Texas Instruments TMS320C40) that were attached to a SUN SPARC host system. The signals were read from hard disk and copied in blocks to the memory of the DSP system. The DSPs carried out the calculations and the results (i.e., the histograms of interaural parameters) were transferred back to the host system and stored on hard disk. The system allows for taking input signals from either hard disk or A/D converters, ensuring real-time processing and calculation of binaural parameters and its distributions. A detailed description of the system can be found in Wittkop *et al.* (1997).

E. d' analysis of differences in interaural parameters

We assume that a classification system discriminates between two directions with a Bayesian criterion on the basis of the observed values of short-term interaural parameters. No further information loss is assumed. In this case and with the further approximation of a Gaussian PDF, the detectability d' of differences in interaural parameters for two different

directions can be calculated as the difference in mean values of the respective distributions divided by the geometric mean of their standard deviation. This measure was calculated from the first moments of the empirical univariate distributions of either ILDs or IPDs, neglecting the deviation from a Gaussian shape.

F. Moment analysis

The marginal PDF estimates were further analyzed by means of descriptive statistics, (i.e., calculation of higher-order moments of the distributions). Following Tukey's rule (Sachs, 1992) for the estimation of the nth moment, at least 5^n samples of a random variable should be taken. Here, at least 1875 samples were evaluated (15-s duration at 125-Hz sampling frequency), so that the first four moments can be estimated according to Tukey's rule. Linear moments, namely expected value, standard deviation, skew, and kurtosis (sometimes also called kurtosis excess), were used for the linear ILD variable. For the distributions of the IPD variable, trigonometric moments as defined by Fisher (1993, p. 41) were used to calculate the mean phase angle, vector strength, circular standard deviation, circular skew, and circular kurtosis (see the Appendix for definitions). By using the trigonometric moments, no unwrapping of the IPDs is required.

G. Bayesian analysis for sound localization

From an information theoretic point of view, the extraction of directional information from the noisy short-term interaural parameters can be described as a Bayesian maximum *a posteriori* (MAP) estimate, which derives the most probable direction based on *a priori* knowledge of the PDF of the parameters for all directions. From Bayes's formula, the conditional probability for each direction λ out of N_λ possible directions is calculated given the parameter vector $\vec{\Delta}$ [cf. Eq. (9)] as

$$P(\lambda|\vec{\Delta}) = \frac{P(\vec{\Delta}|\lambda)P(\lambda)}{\sum_{\lambda=1}^{N_\lambda}[P(\vec{\Delta}|\lambda)P(\lambda)]}, \tag{19}$$

where $P(\vec{\Delta}|\lambda)$ is the conditional probability[2] of $\vec{\Delta}$ given the direction λ and $P(\lambda)$ is the *a priori* probability for the occurrence of the direction λ. From Eq. (10) and under the assumption that all directions λ are equally probable, this formula yields

$$P(\lambda|\vec{\Delta}) = \frac{\prod_{b=1}^{2B} P_b(\Delta_b|\lambda)}{\sum_{\lambda=1}^{N_\lambda}\prod_{b=1}^{2B} P_b(\Delta_b|\lambda)}. \tag{20}$$

The right-hand side arguments are all known, as the $P_b(\Delta_b|\lambda)$ are estimated from the empirical analysis for all b and λ. The marginal distributions $P_b(\Delta_b|\lambda)$ form the *a priori* knowledge used to calculate the probability of all directions. The direction chosen as the most probable, given an observation of the parameter vector $\vec{\Delta}$ and assuming that one source is present in the given noise field, is then

$$\hat{\lambda} = \underset{\lambda\in[1,\ldots,N_\lambda]}{\arg\max}\ P(\lambda|\vec{\Delta}). \tag{21}$$

Using Eqs. (20) and (21), the *a posteriori* probability of all directions can be calculated and the most probable direction can be selected from one observation of parameter $\vec{\Delta}$ and the known distributions of its components for the different directions. A new set of probabilities can be calculated from every time frame so that an ongoing estimate of the most probable direction results.

The statistical localization model described above has been applied to several types of signals and spatially distributed noises. Results for a 25-s segment of speech from a single talker (different from the four-talker target signal) in cafeteria noise at 5 dB SNR are reported below. The distributions of the four-talker target and the cafeteria noise sample at 5 dB SNR were used as reference distributions[3] in Eq. (20).

The *a posteriori* probabilities were calculated for the total signal length, yielding 3200 samples of the probabilities for each of the 430 probed directions. The *a posteriori* probabilities were smoothed by a first-order, low-pass filter with 100-ms time constant, and the most probable direction was determined from the smoothed probabilities. The estimates for the most probable direction are plotted into a normalized histogram, which describes the probability that a specific direction is detected as the most probable direction, given the real direction. This plot can be described as a "decision histogram," displaying deviations and confusions in the estimates in one view.

III. RESULTS

A. Distributions of interaural parameters

In this section, the empirical results on the statistics of short-term interaural parameters are described. Specifically, the dependence of the distributions on frequency, SNR, direction, and noise type is analyzed. Due to the large amount of data, the parameter space covered in this analysis had to be restricted. The signal conditions pertaining to speech in silence and speech in cafeteria noise at a moderate-to-low SNR of 5 dB were used as primary examples for one lateral direction and one direction near the median plane (0° elevation). Results for low frequencies are reported for the IPD variable (340 and 540 Hz), and results for medium and high frequencies are reported for the ILD variable (830 Hz and 2.88 kHz).

1. Dependence on frequency

Figure 2 shows percentiles of the distribution of the ILD variable as a function of frequency for the different conditions pertaining to speech in silence (upper panel), speech in cafeteria noise at 5 dB SNR (middle panel), and car interior noise at 5 dB SNR (lower panel).

In each condition, the 5, 10, 25, 50, 75, 90, and 95 percentiles of the distributions are plotted for −15° azimuth (lines) and +85° azimuth (symbols) and 0° elevation. The widths of the distributions can be assessed, e.g., by considering the vertical difference between the 95 percentile line and the 5 percentile line. For the case of silence, the widths of the distributions reveal that the parameter fluctuation is considerable, even though no additional noise is present,

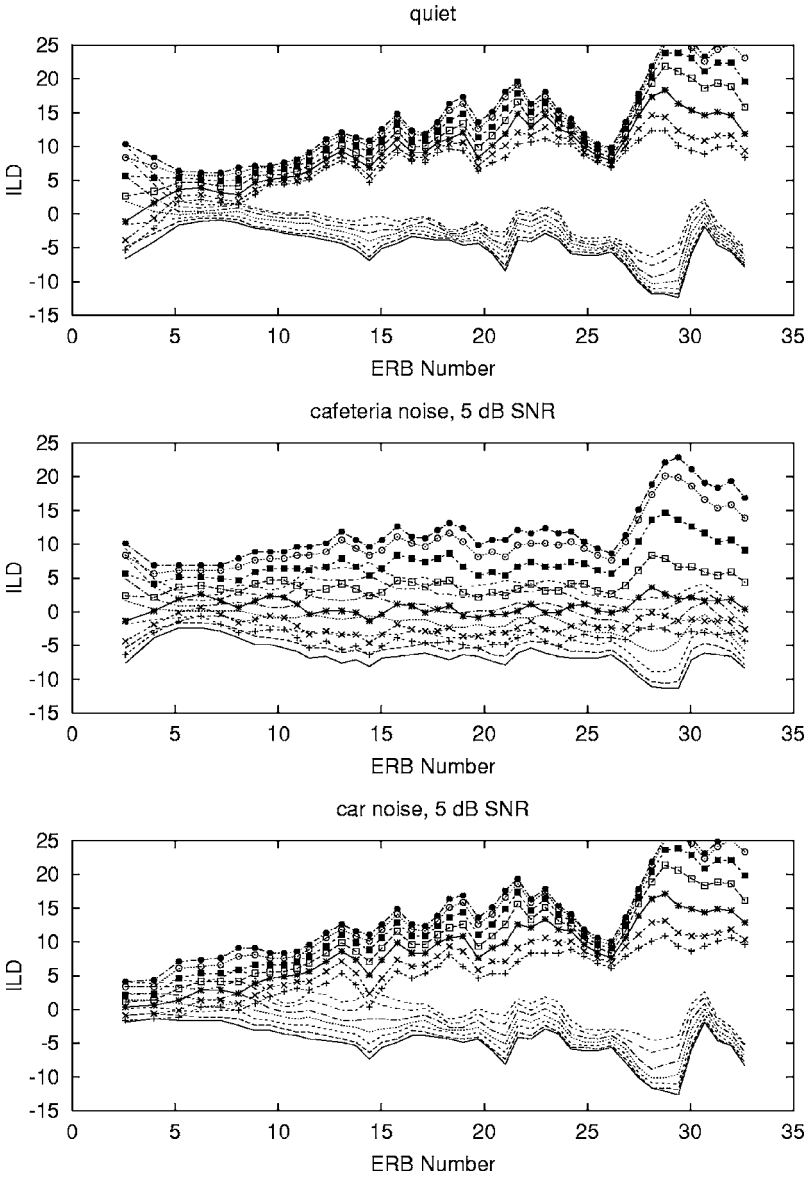

FIG. 2. Distribution percentiles for the interaural level difference (ILD) as a function of frequency for three different conditions: speech target in silence (upper panel), in cafeteria noise (5 dB SNR, mid panel), and in car interior noise (5 dB SNR, lower panel). Each panel shows 5, 10, 25, 50, 75, 90, and 95 percentiles for −15° azimuth (symbols) and +85° azimuth (lines) and 0° elevation. Percentiles are calculated by integration from the negative towards positive parameter values.

which is due to the statistics of the signal in short time frames. The width depends on both the frequency and the direction of the target. It is especially high in the lowest two frequency bands (50 and 100 Hz) because the band level of the speech is close to the recording noise level in these bands. The distributions are narrow as compared to the mean difference between these directions, except for the low frequencies of up to 8 ERB (310 Hz). In the noise conditions, however, the distributions are significantly broadened so that they overlap for the different directions. The broadening extends across the whole frequency range in the cafeteria-noise condition because the long-term frequency spectra of target and noise are nearly the same and the frequency-specific SNR is almost constant. In the car interior noise, the broadening is restricted to the lower frequency region, because the noise has a $1/f$ type of spectrum with primarily low-frequency content. The SNR of the remaining frequencies is higher than for the cafeteria noise at the same overall SNR.

2. Dependence on signal-to-noise ratio (SNR)

In order to clarify the influence of noise on the distributions of interaural parameters, their dependence on SNR was studied. Figures 3 and 4 show the distributions of the ILDs and IPDs, respectively, for the speech target in cafeteria noise. Distributions are plotted for two directions (15° and 60° azimuth 0° elevation) and two frequencies (IPDs: 340 and 540 Hz; ILDs: 830 Hz and 2.88 kHz). For the ILDs, each of the four plots shows distributions for the SNR values of −5, −2, −1, 0, 1, 2, 3, 5, 10, 15, 20, and 30 dB, and in silence. The curves are separated for clarity by relative shifting in y direction by 0.025 (ILDs), the uppermost curve in each plot showing the distributions in silence.

In silence, the ILD distributions (Fig. 3) are relatively narrow and show little overlap between directions at both frequencies (upper versus lower panels, respectively). The ILD increases significantly with frequency for both directions (left versus right panels, respectively). It is therefore a distinctive parameter for the direction. However, the distributions "decay" with decreasing SNR due to the influence of the noise. Specifically, the variance increases and the mean value is shifted towards zero. The distributions are skewed at medium SNRs. These effects are due to the nonlinear superposition of the PDFs of the target, which has a nonzero mean

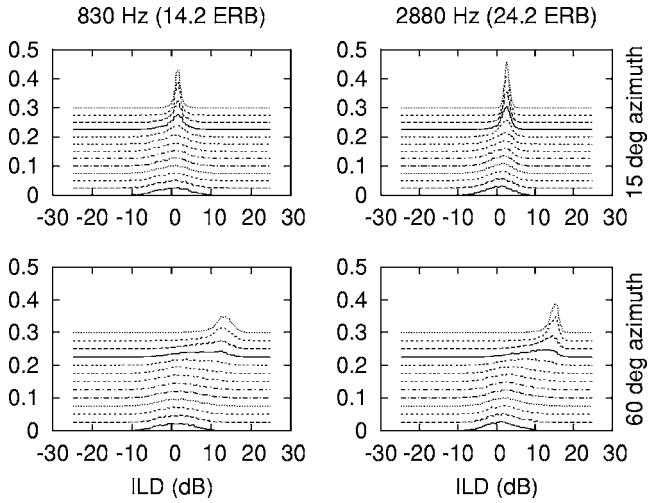

FIG. 3. Histograms of ILD values, i.e., number of observations of a specific ILD normalized to the total count, for the speech target in cafeteria noise. The left panels are for a frequency of 830 Hz and the right panels are for 2.88 kHz. The upper panels are for 15° azimuth and the lower panels are for 60° azimuth (0° elevation, respectively). Each panel shows the distributions for the SNR values of −5, −2, −1, 0, 1, 2, 3, 5, 10, 15, 20, and 30 dB and in silence. The curves are shifted in this order successively by 0.025 in the y direction for clarity.

and a low variance due to its directionality, and of the noise, which has a zero mean and higher variance because of its diffusiveness. Both the increased variance and the systematic shift of the mean value towards zero lead to a reduction of the variation of the distributions with direction and frequency. The systematic shift is especially large for the lateral direction of 60° azimuth (lower panels). In this case, the observed level difference at low SNRs approaches the difference between the noise level at the contralateral ear and the source level at the ipsilateral ear rather than the large level difference expected from the anechoic HRTFs.

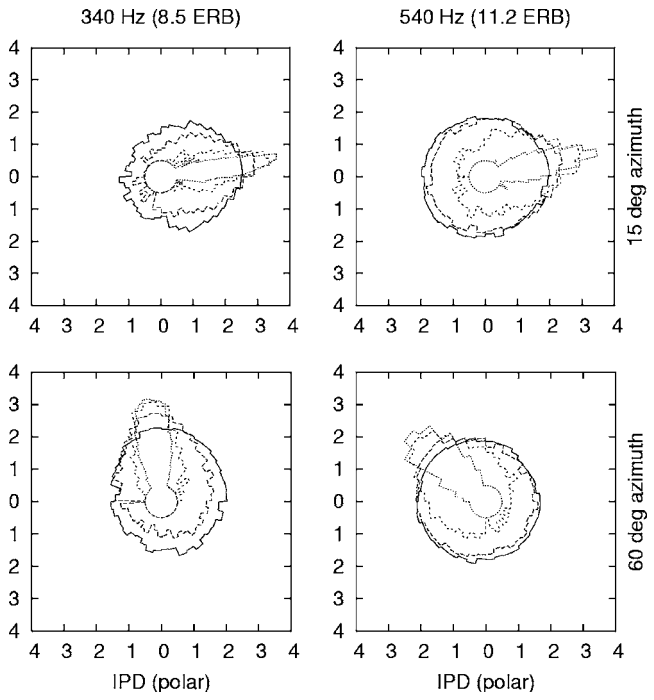

FIG. 4. Polar plots showing histograms of the interaural phase difference (IPD). The frequency subbands are 340 Hz (left panels) and 540 Hz (right panels), the azimuths 15° (upper panels) and 60° (lower panels); the elevation was 0°. The angle parameter of the plot shows the angle of the IPD, in counter-clockwise orientation, and the angle 0° corresponds to the half-axis with $y=0$, $x>0$. The radius parameter of the curve, r, shows the relative frequency of occurrence h of this IPD value. To make the differences visible, the frequencies of occurrence are logarithmically transformed according to $r=4.5+\log_{10} h$. The maximum radii of each curve are ordered as the SNR values at which the histograms were measured. The SNRs are, in the order from the largest maximum to the smallest, silence, 20, 5, and −5 dB.

Figure 4 shows the corresponding histograms of the IPD variable; because the variable is cyclical, the histogram is plotted in a polar diagram. The angular parameter of the plot shows the IPD, in counter-clockwise orientation, and the IPD 0° corresponds to the half-axis with $y=0$, $x>0$. The radius of the curve, r, shows the relative frequency of occurrence h of this IPD value. To make the differences visible, the frequencies of occurrence are logarithmically transformed according to $r=4.5+\log_{10} h$. The curves are directly superposed without offset. The maximum radii of each curve are ordered as the SNR values in which the histograms were measured. The figure shows that for the IPD variable the PDFs "decay" as well, converging to a uniform circular distribution with decreasing SNR. The variation of the mean value with decreasing SNR is much smaller than for the ILD variable.

The higher-order moments of the ILD distributions at 830 Hz and 2.88 kHz and 60° azimuth (lower left and right panel in Fig. 3) are listed in Table II. Both frequencies behave similarly. The mean values and standard deviations show the broadening and shifting described above. Due to the nonlinear superposition of the PDFs of target and noise, the standard deviations show a nonmonotonic behavior. They increase and then decrease slightly with decreasing SNRs. Skew and kurtosis are significantly different from 0 at high SNRs and converge towards zero with decreasing SNRs. Again, nonmonotonic behavior is observed with decreasing SNRs.

The corresponding higher-order moments of the IPD distributions for the frequencies 340 and 540 Hz are shown in Table III. The mean values depend less on the SNR, whereas the standard deviations show a monotonic increase with decreasing SNRs at both frequencies. Nonzero skew and kurtosis values are observed especially at the lower frequency. For a Gaussian PDF, both parameters would be zero because a nonzero skew indicates an asymmetrical distribution and a positive kurtosis indicates a distribution with a sharper maximum and wider shoulders than the normal distribution. Here, the skew varies between 0.3 and 2.0, and the kurtosis has values up to 30.9, which is a very large difference in shape from Gaussian PDFs.

The analysis of higher-order moments shows that non-Gaussian PDFs have to be assumed in general for the distributions of the interaural parameters. Whether the small deviations from Gaussian shape at low SNRs are still relevant for the retrieval of binaural information remains to be further analyzed.

3. Dependence on direction

The HRTF-derived interaural parameters of stationary, undisturbed signals show a clear dependence on direction,

TABLE II. Moments derived from the distributions of the ILD variable at 830 Hz (left columns) and 2.88 kHz (right columns) at various SNR. The target-noise condition was speech at 60° azimuth and 0° elevation in cafeteria noise. Moments are $\bar{x}$ expected value, σ standard deviation, s skew, and K kurtosis.

		f=830 Hz				f=2880 Hz			
Noise type	SNR (dB)	$\bar{x}$(dB)	σ(dB)	s(1)	K(1)	$\bar{x}$(dB)	σ(dB)	s(1)	K(1)
cafeteria	silence	12.83	2.52	−0.92	1.78	14.63	1.68	−2.03	8.04
cafeteria	30	11.68	3.59	−1.26	2.35	14.09	2.15	−1.63	4.09
cafeteria	20	8.82	5.05	−0.70	0.12	11.78	3.81	−1.08	1.20
cafeteria	15	6.82	5.46	−0.31	−0.48	9.64	4.56	−0.58	−0.31
cafeteria	12	5.61	5.53	−0.12	−0.54	8.08	4.84	−0.33	−0.53
cafeteria	10	4.83	5.51	−0.02	−0.52	7.02	4.93	−0.17	−0.55
cafeteria	5	3.13	5.21	0.10	−0.21	4.49	4.83	0.13	−0.27
cafeteria	3	2.59	5.03	0.11	−0.21	3.63	4.66	0.21	−0.11
cafeteria	2	2.32	4.94	0.10	−0.17	3.22	4.55	0.24	−0.01
cafeteria	1	2.08	4.86	0.08	−0.15	2.82	4.50	0.17	0.24
cafeteria	0	1.87	4.77	0.08	−0.13	2.52	4.36	0.25	0.09
cafeteria	−1	1.67	4.69	0.05	−0.13	2.21	4.27	0.23	0.12
cafeteria	−2	1.49	4.61	0.04	−0.12	1.93	4.17	0.21	0.11
cafeteria	−5	1.04	4.42	−0.01	−0.11	1.24	3.93	0.11	0.05

and extensive sets of data exist on this in the literature (e.g., Wightman and Kistler, 1989a). Therefore, the direction dependence of the expected values of the short-term interaural parameters, which are associated with the HRTF-derived parameters, is not shown here. Instead, higher-order moments of the distributions are considered.

Data not shown here reveal that the ILD standard deviations depend strongly on the azimuth (for 15 dB SNR, they vary from about 2 to 5 dB at azimuth angles from 15° to 90°) and only moderately on the elevation (about 1 dB variation from −20° to 45° elevation at 15 dB SNR). Standard deviations are high for lateral directions, where the ILDs themselves are large. For the IPD variable, however, the vector strength is high for the lateral directions (i.e., for this frequency, the short-term phase vectors are more congruent in time than for the more central directions). For an SNR of 5 dB, the variation in the vector strength is from about 0.18 to 0.52 for a variation in azimuth from 15° to 90° and for a variation in elevation from −20° to 45°. It is clear from this analysis of the second moments that the fluctuation of the short-term interaural parameters depends strongly on the direction at a fixed SNR. IPDs and ILDs behave differently in this aspect.

4. Dependence on noise condition

The moments of the distribution of the ILD variable at 830 Hz and 2.88 kHz and of the IPD distributions at 340 and 540 Hz for various target-noise conditions are listed in Table IV. The target was always speech at 60° azimuth and 0° elevation and the SNR was 15 dB. The moments for the speech target in silence are included as a reference.

The data for the ILD show that the parameter values lie in a narrow range as compared to the deviation from the

TABLE III. Same as Table II, but for the IPD variable at 340 and 540 Hz. The moments are ϕ expected value of phase angle, σ_z standard deviation, s_z skew, and K_z kurtosis (see the Appendix for definitions).

	f=340 Hz				f=540 Hz			
SNR (dB)	ϕ(rad)	σ_z(rad)	s_z(1)	K_z(1)	ϕ(rad)	σ_z(rad)	s_z(1)	K_z(1)
silence	1.69	0.12	0.43	4.77	2.40	0.17	0.74	21.11
30	1.69	0.14	0.29	10.56	2.42	0.32	−0.05	23.74
20	1.67	0.26	2.00	30.92	2.46	0.60	−0.39	6.69
15	1.65	0.38	1.59	17.84	2.51	0.82	−0.23	2.99
12	1.63	0.48	1.24	11.79	2.55	0.96	−0.18	1.77
10	1.62	0.55	1.06	8.57	2.58	1.05	−0.16	1.27
5	1.54	0.77	0.62	3.64	2.71	1.28	−0.13	0.53
3	1.50	0.86	0.45	2.46	2.77	1.37	−0.10	0.35
2	1.47	0.91	0.44	1.99	2.79	1.41	−0.10	0.29
1	1.43	0.96	0.39	1.61	2.82	1.45	−0.09	0.23
0	1.39	1.02	0.36	1.27	2.85	1.48	−0.09	0.20
−1	1.35	1.07	0.32	1.02	2.88	1.51	−0.07	0.17
−2	1.29	1.13	0.30	0.80	2.92	1.54	−0.06	0.14
−5	1.08	1.28	0.24	0.36	3.03	1.61	−0.03	0.09

TABLE IV. Distribution moments of the ILD variable at 830 Hz (left table entry) and 2.88 kHz (right table entry) and for the IPD variable at 340 and 540 Hz for various target-noise conditions. The target was always speech at 60° azimuth and 0° elevation and the SNR was 15 dB. The "speech in silence" condition is listed as a reference. The noise condition "inside car" is listed separately, the data for all other noise conditions listed in Table I were averaged and the mean and standard deviation are given.

		f=830 Hz				f=2880 Hz			
Noise type ILD	SNR (dB)	$\bar{x}$(dB)	σ(dB)	$s(1)$	$K(1)$	$\bar{x}$(dB)	σ(dB)	$s(1)$	$K(1)$
—	silence	12.83	2.52	−0.92	1.78	14.63	1.68	−2.03	8.04
mean	15	6.64	5.55	−0.37	−0.34	9.56	4.47	−0.65	0.24
(std)	15	(2.07)	(0.73)	(0.20)	(0.21)	(0.84)	(0.50)	(0.15)	(0.48)
inside car	15	12.43	2.95	−0.84	2.67	14.58	1.83	−1.03	14.03
		f=340 Hz				f=540 Hz			
IPD	(dB)	$\bar{x}_z$(rad)	σ_z(rad)	$s_z(1)$	$K_z(1)$	$\bar{x}_z$(rad)	σ_z(rad)	$s_z(1)$	$K_z(1)$
—	silence	1.69	0.12	0.43	4.77	2.40	0.17	0.74	21.11
mean	15	1.67	0.29	1.18	20.91	2.45	0.59	−0.15	8.10
(std)		(0.02)	(0.08)	(0.81)	(5.38)	(0.07)	(0.15)	(0.88)	(4.57)
inside car	15	1.68	0.18	2.10	25.14	2.41	0.21	0.92	25.47

values in the silent condition. This shows that the influence of the noise type on the distributions is small relative to the influence of the SNR (cf. Table II). The only significant deviation is observed for the car interior noise, where the parameters resemble those of the silent condition. The mean value for the IPD variable is similar in all noise conditions and deviates little from the silent condition. For the standard deviation, an increase is observed, which is larger at the higher frequency. The skew and kurtosis, however, vary across noise conditions more than for the ILDs. Nevertheless, similar noise types have a similar impact on these parameters.

It can be concluded from the analysis of the moments of the distributions that different types of spatially distributed noise have a similar influence on the distribution of short-term interaural parameters. The SNR is therefore the most relevant parameter for the quantification of the noise's impact on mean and variance. However, the higher-order moments (i.e., skew and kurtosis) vary with the noise condition, especially for the IPD variable.

B. Simulation results

A probabilistic approach of directional information extraction from short-term interaural parameters is studied in this section. Both discrimination of directions and absolute localization are considered.

1. d' analysis of differences in interaural parameters

Figure 5 shows single-band d' derived from two different target directions as a function of SNR and in silence for the speech in cafeteria noise condition. Data are plotted for the ILD variable at 830 Hz (+) and 2.88 kHz (□) and for the IPD variable at 340 Hz (×) and 540 Hz (∗).

In the upper panel of Fig. 5, the two target directions are 0° and 5° azimuth in the horizontal plane. In silence, d' is greater than 1 in all cases except for the ILD variable at 830 Hz. However, d' decreases significantly with decreasing SNRs and reaches a value of about 0.2 on average at an SNR of 5 dB. Assuming that the observations in different frequency bands are statistically independent, and that the ILDs vary systematically with small azimuth changes, the d' increases with a factor of the square root of the number of observations. The number of observations required to detect a difference in direction (i.e., $d'=1$) is about 25 in this case.

In the lower panel of Fig. 5, the differences in interaural parameters arise from a shift in elevation from 0° to 15° in the median plane. The d' is lower than in the case of azimuth

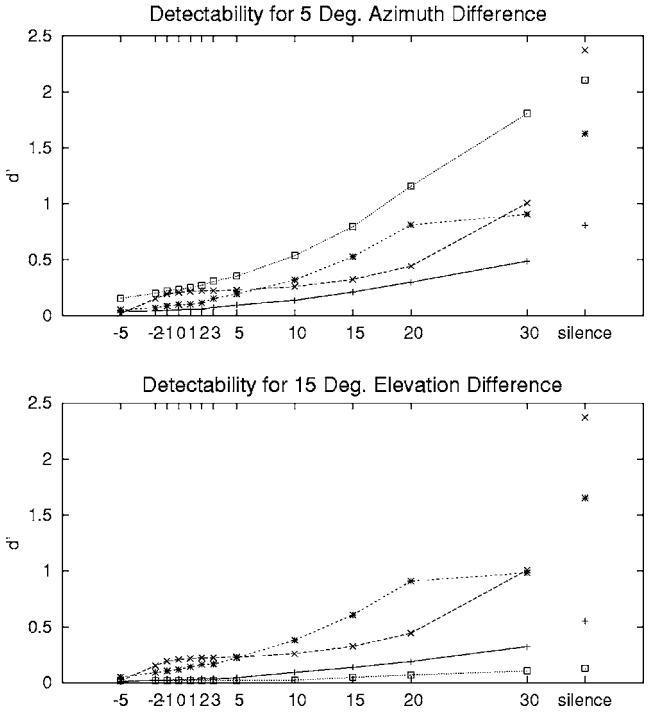

FIG. 5. d' of differences in interaural parameters derived from two different target directions as a function of SNR. In the upper panel, the two target directions were at 0° and 5° azimuth in the horizontal plane. In the lower panel, the differences in interaural parameters arise from a shift in elevation from 0° to 15° in the median plane. The target-noise condition was speech in cafeteria noise. Each plot shows data for the ILD variable at 830 Hz (+) and 2.88 kHz (□), and for the IPD variable at 340 (×) and 540 Hz (∗).

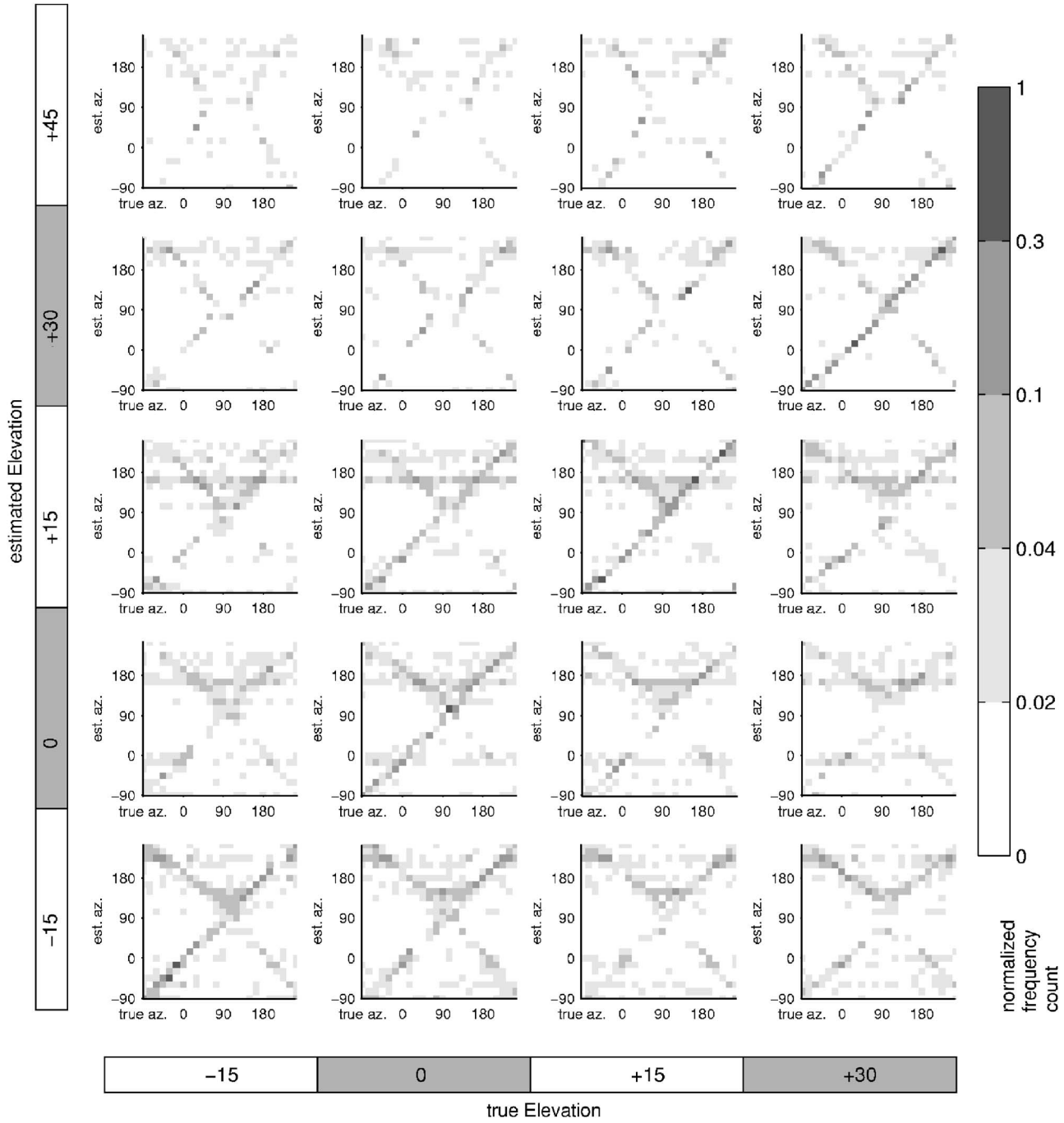

FIG. 6. Decision histogram for the target-noise condition "speech in cafeteria noise" at 5 dB SNR. The *y* axis represents the detected direction and the *x* axis the real direction of the target. Plotted is the normalized frequency count of localization decisions as an estimate of the probability that a specific direction is detected as the most probable direction, given the real direction. Each box in the plot gives real and estimated azimuth for a given combination of real and estimated elevation. If all decisions are correct, a diagonal intersecting the lower left plot's origin should result.

variation. At 5 dB SNR, d' is on average 0.1. In this case, about 100 independent observations, combined across time, frequency, or both, are needed to reach a d' of 1.

2. Simulation of absolute sound localization

Figure 6 gives the decision histogram in columns for each direction as a gray-scale-coded frequency count. The condition is speech in cafeteria noise at 5 dB SNR. Each small rectangle in the plot represents a real and estimated azimuth for a given combination of real and estimated elevation. Real azimuths are varied along the *x* axis of the subplots, estimated azimuths along the *y* axis of the subplot. Each real elevation is represented in a column of subplots and each estimated elevation in a row of subplots. If all decisions are correct, a black diagonal intersecting the plot's origin should result. Decisions plotted on parallel lines intersecting the *y* axis at different elevation boxes signify elevation confusions, whereas perpendicular diagonal lines indicate front-back confusions. Most pixels away from the diagonal are white, indicating that less than 2% of the direction estimates were given for this real/estimated direction combination. If for a fixed "true" direction estimates would

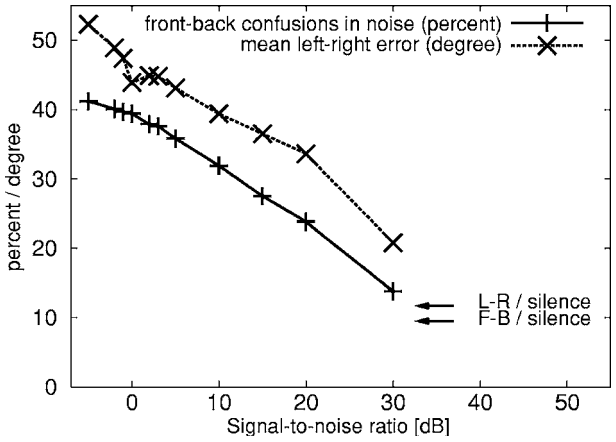

FIG. 7. Percentage of front-back confusions (+) and rms error of the angle to the median plane (×) of the Bayes localization simulation as function of SNR. The target-noise condition is speech in cafeteria noise. Data were averaged for all target directions. An estimate was considered a front-back confusion if the total angle of error was decreased by at least 15° when mirroring the azimuth coordinate at the frontal plane.

be evenly distributed across all possible directions, a white column across all subplots would result. The top row of the subplots is mostly white because the elevation value of 45° did not occur in the test data.

The percentage of front-back confusions, averaged across all target directions, was calculated from the decision histogram for the target-noise condition "speech in cafeteria noise" as a function of the SNR. A directional estimate was defined as a front-back confusion if the total angle of error was decreased by at least 15° when mirroring the azimuth coordinate at the frontal plane. Additionally, the direction estimates and the true directions were transformed to the angle to the median plane, also known as the "left-right" coordinate (Wightman and Kistler, 1989b), and the rms value of the error across all tested directions was evaluated. Figure 7 shows the percentage of confusions (+) and the rms error of the angle to the median plane (×). The percentage of confusions increases from 9.5% in silent conditions to 41.2% at the lowest SNR of −5 dB, while the rms value of the left-right coordinate has a value of 11.7% in silent conditions, which rises to 52.3% at −5 dB SNR.

For a real-world application of the Bayes localization algorithm, its robustness against changes in the noise environment is crucial. The dependence of the distributions of interaural parameters on the SNR was found to be especially high so that the localization accuracy might be lowered, if the reference distributions used in Eq. (20) did not match the test stimuli in SNR. The localization accuracy for the target-noise condition "speech in cafeteria noise" has therefore been evaluated for different SNRs with *a priori* information measured at the same SNR ("matched" condition) and at different SNRs ("unmatched" condition). Data are shown in Table V.

The rows show the percentage of front-back confusions for the different test SNRs. The columns give the SNR of the reference distributions that were used as *a priori* knowledge. The conditions with optimal information form the diagonal of the matrix (same data as in Fig. 7). For a test SNR of 10 dB (fifth row), the percentage is lowest for the optimal condition, i.e., a reference SNR of 10 dB. However, the percentage is only slightly higher for an unmatched reference SNR. With a fixed reference SNR of 15 dB (fourth column), the dependence on the test SNR is similar to the dependence with matched references (entries on the diagonal). Usually, very small differences among data in the same row are observed. However, if the reference condition which was measured in silence is tested at medium SNRs, the percentage of confusions increases from 9.5% to 34.2% (first row).

IV. DISCUSSION

A. Consequences of parameter distributions for sound localization

The analysis described in the last section reveals significant variations of short-term interaural parameters in silent environments and especially in noise conditions, justifying their statistical description as random variables. The empirical approach of estimating probability distributions from observations of ILD and IPD time series from actual binaural recordings seems to be sufficient for the characterization of random variables, because it reveals the most relevant dependencies of their PDFs on the sound field properties.

The description of short-term interaural parameters as random variables has several consequences for the extraction

TABLE V. Percentage of front-back confusions for the Bayes localization algorithm for the target-noise condition "speech in cafeteria noise." Rows show the data for the different tested SNR. The columns give the SNR of the reference distributions that were used as *a priori* knowledge. The conditions with optimal information form the diagonal of the matrix.

	Silence	30	20	15	10	5	3	2	0	−2	−5
Silence	9.5			34.2							
30		13.8		11.4							
20			23.8	22.3							
15				27.5							
10	34.2	33.2	33.6	32.9	31.9	32.0	32.4	31.8	35.7	34.7	35.0
5				37.1		35.8					
3				37.9			37.6				
2				39.0				37.9			
0				39.6					39.4		
−2				40.1						40.1	
−5				41.4							41.2

of directional information from short-term interaural parameters. First, the parameters fluctuate because of the nonstationarity of the signals and their nonlinear combination. Therefore, information retrieval at low SNRs can only be performed in a statistical sense and requires a certain number of observations of short-term interaural parameter values. Second, directional information can be retrieved with *a priori* knowledge of the statistics of the interaural parameters. Restriction of the *a priori* knowledge to the interaural parameters derived from anechoic HRTFs can lead to much larger errors, as Table V shows. Third, the detectability of systematic variations of interaural parameters with direction might be reduced due to their fluctuations. The amount of fluctuations is mainly a function of SNR and source direction and in the examined cases does not depend much on the noise type so that a general quantification of this effect as a function of SNR seems appropriate. Fourth, the systematic variation of the ILDs with direction itself is reduced due to the shift in mean values, i.e., the bias. Due to the SNR-dependent bias, it can be assumed that the large and significant ILD values observed in anechoic HRTFs from lateral directions cannot be fully exploited for localization or discrimination of directional sources in noise without taking into account the SNR. Finally, the direction dependence of higher-order moments possibly has to be taken into account. It could in principle be exploited for direction estimation, if an analysis of higher-order statistics is included.

B. Detectability of differences in interaural parameters

The d' analysis shows that at moderate-to-low SNRs of about 15 dB or below, the physical difference in interaural parameters induced by significant variations of direction, according to our assumptions, is not detectable on the basis of the observation of either one of the variables alone. However, discrimination is possible by integration of the information across frequency, time, or both. The coarse estimate given above shows that the number of observations necessary for the detection of 5° azimuth or 15° elevation difference is in the range of the number of frequency bands in a critical band analysis of interaural parameters. In order to reach the same number of observations by temporal integration, time windows of about 400 to 800 ms would be needed, which are much larger than psychoacoustically derived time constants of localization in humans (Stern and Bachorski, 1983). Some authors discussed that interaural cues caused by asymmetries between the pinnae may be sufficient also for sound source localization along the "cones of confusion" and the median plane (Searle *et al.*, 1975). However, as these asymmetries provide only small cues, it still has to be understood in which situations they are relevant.

If such an integration mechanism is used by the human hearing system, it can be assumed that the physically defined d' is perceptually relevant, because at moderate-to-low SNRs, the standard deviation of the IPDs at low frequencies and of the ILDs at medium-to-high frequencies is larger than the human just-noticeable differences (jnd's) in ILDs and IPDs in the respective frequency regions (cf. Ito *et al.*, 1982; Stern *et al.*, 1983). It can thus be assumed that the external noise is the limiting factor in this range of SNRs rather than internal noise associated with neural processing.

The notion that frequency integration is needed to ensure detectability, although physically founded here, coincides with the physiological studies by Brainard *et al.* (1992) in barn owls that demonstrated binaural information processing in multiple frequency bands and subsequent frequency integration. Directional ambiguities have been shown to be resolved by this integration. The assumption of frequency integration is supported by psychophysical findings showing that the localization accuracy increases with increasing bandwidth of the stimulus (e.g., Butler, 1986).

C. Simulation of absolute localization

Figure 6 shows that the confusion pattern is qualitatively similar to the patterns of human performance found, e.g., by Good and Gilkey (1996). Confusions occur mainly as elevation errors or front-back confusions. Hardly any confusions occur that could not be explained by this typical pattern. At lower SNRs, localization of lateral directions becomes more blurred, which is not shown here.

Good and Gilkey obtained human data for a single noise masker in the front and a click train from different directions. The SNR was defined relative to the detection threshold of the target for the condition where the target was in the front, other directions were tested with the same free field sound level. That means that the SNR at the ear canal entrance varied for different directions of the sound source. Due to the different target-noise condition and the different definition of the SNR, the results are not directly comparable. The diffuse cafeteria condition with a single talker target is probably more difficult to localize at the same SNR than the single-noise source condition using a click train as a target employed for the psychophysical experiments. Also, the psychoacoustic experiment was performed at a constant free-field SNR for each trial block. Because human pinnae enhance the high-frequency part of the spectrum of sounds from frontal directions (Shaw, 1997), there is a systematic direction dependence of the SNR at the ear canal entrance, so that detectability may have been used by the subjects as a cue to estimate the sound source azimuth. A more thorough comparison of the confusion patterns of the algorithm and in humans using the same stimulus configuration is therefore indicated. However, it can be concluded from the data illustrated here that the localization accuracy of the model is qualitatively similar to the one in humans. The relatively good estimation of elevations and front-back directions in the median plane with binaural input only can be explained by the fact that the algorithm is able to exploit asymmetries of the pinnae by the across-frequency integration of probabilities.

Table V suggests that the performance of the algorithm depends mainly on the SNR of the test condition and only slightly on the reference SNR. This finding shows that the Bayes localization model is robust against changes of the SNR of the reference condition. Its performance decreases only slightly if the *a priori* knowledge does not match the actual target-noise condition to be analyzed. The results

show that the distribution of interaural parameters as measured here could be one possible robust source of *a priori* information.

There are several reasons why human subjects probably are able to use binaural information in a more efficient way. First, the proposed MAP estimator disregards possible correlations between frequency channels, because in Eq. (10) the multidimensional PDF is approximated as a product of its marginals. Jenison (2000) has shown that a maximum-likelihood estimator with knowledge of the response covariance structure is able to perform better on a correlated population response than an estimator assuming independence. This might be relevant here, because the auditory system effectively possesses not only 43 frequency channels, but many thousands of nerve fibers with overlapping receptive fields. Second, the measured distributions include small variations of the interaural parameters due to head movements during the recording. These measurement errors decrease the localization performance at high SNRs. Third, humans use frequencies of at least up to 10 kHz for sound localization, and the high-frequency ILDs are probably particularly important, while the frequencies used here do not exceed 8 kHz. Fourth, the simulation does not include the interaural group delay, corresponding to time differences of the envelopes, which can be computed, e.g., according to Eq. (4). So far, the role of high-frequency ITDs has not been clarified completely (Macpherson and Middlebrooks, 2002); for high frequencies, envelope delays are probably more important. Therefore, it is possible that including some representation of interaural envelope delays improves localization performance. Fifth, humans can use monaural cues for sound localization in some circumstances. This can especially improve discrimination of directions on the median plane. Because monaural cues are not evaluated in the simulation, the distinction of front-back directions and elevations along the "cone of confusion" is probably worse than the performance of humans at the same SNR. Sixth, the used resolution of the histograms of IPDs is in part of the cases coarser than psychoacoustically observed jnd's; this should only affect localization at high SNRs. On the other hand, there is one aspect which may improve the performance of the algorithm as compared to human subjects, i.e., that humans cannot extract the fine structure of waveforms beyond 2 kHz. However, at higher frequencies, the IPDs are not only strongly disturbed by noise, but also become highly ambiguous. Taking all preceding aspects into account, humans probably can use the binaural information in a more efficient way, especially for directions close to the median plane, and for higher frequencies. A preprocessing model which better matches human binaural processing including interaural envelope delays, and excluding IPDs for channels at 1.5 kHz and higher, can possibly explain most of the localization ability of humans by binaural parameters alone.

It should be noted that the Bayesian approach is equivalent to the one of Duda (1997) if one assumes that the distributions are Gaussian with constant variance. In this case, the MAP procedure reduces to a least-squares fit, which would not need the *a priori* knowledge of all distributions. In contrast, Eq. (20) allows a more general approach which is able to take noise explicitly into account.

D. Frequency integration of probabilities

Most models of lateralization or localization combine short-term correlation values over frequency by a summation or multiplication (e.g., Stern *et al.*, 1988; Shackleton *et al.*, 1992; Stern and Trahiotis, 1997; Braasch and Hartung, 2002; Braasch, 2002b, a). Neurophysiological findings support that, for some species, after the detection of interaural parameters, a frequency integration is performed. This has been shown by Brainard *et al.* (1992) in the barn owl. Interaural cue detection followed by frequency integration has been used also successfully by frequency-domain models and technically motivated algorithms of sound localization (Duda, 1997; Wittkop *et al.*, 1997; Nakashima *et al.*, 2003). In contrast to the models cited above, the quantities which are integrated across frequency in the model presented here are *probabilities*, which takes, according to the assumptions stated, the available information into account in an optimum way.

E. Statistical representation of interaural timing

Because interaural parameters are considered in the frequency domain in narrow frequency bands, the interaural phase differences (IPDs) are used to describe timing differences. IPDs have several advantages over ITDs: For signals filtered by narrow-band auditory filters, the interaural cross-correlation function (ICCF) becomes nearly periodic. In the case that the signal is nonstationary or contains additional noise, the maximum of the ICCF, which is used frequently to estimate the ITD, is not well defined (Lyon, 1983; Stern *et al.*, 1988; Schauer *et al.*, 2000). This ambiguity of the maximum of the ICCF is directly related to the ambiguity of the phase in the frequency domain. This can be explained by the fact that, according to Eq. (8), both representations are linked by the combination of the generalized Wiener-Khintchine theorem (cross-correlation theorem) and the Wiener-Lee relation, and therefore have equal information content. Consequently, the probability density function (PDF) of ITD estimates based on the ICCF would have several maxima. While it is possible to describe such multimodal PDFs by histograms, there is no well-established approach to characterize it by a few parameters.

Contrarily, the statistics of the IPDs can be described neatly by statistics of cyclical data as defined by Fisher (1993); the expected value and variance can be calculated robustly, and empirically observed PDFs can be approximated well by the von Mises distribution. The phase difference is represented in the complex plane. Therefore, the error-prone operation of unwrapping the phase of noisy signals (Tribolet, 1977) is not necessary. This advantage of the IPDs has shown to become especially important in noise, as demonstrated by technical algorithms for robust sound localization (Liu *et al.*, 2000; Nakashima *et al.*, 2003).

F. Possible physiological representations of interaural timing

The processing structure sketched here aims to be a possible description of important features of the binaural auditory system. Clearly, the actual signal processing in binaural processing of interaural timing is still being discussed and may vary between different species. However, the explicit consideration of external noise, as proposed here, might be relevant for modeling physiological data.

Harper and McAlpine (2004), e.g., showed that when assuming a population code for distributions of IPDs for humans, as measured in indoor and outdoor sound fields, there are consequences for optimum distributions of best IPDs of auditory nerve fibers. Fitzpatrick *et al.* (1997) propose a population code based on the observation that localization of sounds is much more accurate than the spatial sensitivity of single neurons. Population codes have been proposed also, e.g., by Hancock and Delgutte (2004). The statistical data as well as the Bayesian approach described here could help to develop such models further, and eventually to decide which model matches neural data best.

The approach described here is solely based on the physical properties of the interaural parameters and subsequent Bayesian estimation. However, there is an interesting similarity with physiological models and data. When taking the logarithm of Eq. (20), the log probability $\log P(\lambda|\vec{\Delta})$ can be interpreted as an activity that is a sum of the activities (log probabilities) derived from the frequency bands. Frequency-specific activities are generated by the log distributions $\log P_b(\Delta_b|\lambda)$ from the observed parameter Δ_b. In terms of neural processing, the log distributions can be interpreted as optimum tuning curves of neurons sensitive to single-channel ILDs and IPDs. These narrow-band tuning curves are rather broad due to the external noise. The tuning curve sharpens by summation across frequency, which is equivalent to multiplication of probabilities, and a precise and robust localization is possible, although the basic tuning curves are unspecific. Measuring the "response" for ITD of narrow-band stimuli would yield periodic tuning curves. Also, with progressive frequency integration, ITD tuning curves would become less periodic and their shape should become more similar to a wideband cross-correlation function, the bandwidth corresponding to the bandwidth of the receptive fields. Therefore, the width and shape of the tuning curves could be interpreted as useful to increase the robustness in noise. The observed deviations of the parameter distributions from Gaussian shapes suggest that properties of physiologically observed tuning curves for interaural parameters, such as asymmetry (skewedness), might be an adaptation to increase the robustness with real-world stimuli.

The interpretation of the log distributions as activation tuning curves and well-defined *a priori* information can be regarded as a major advantage of the Bayesian localization model compared to other approaches that use neural nets in combination with common training rules (e.g., Neti *et al.*, 1992; Datum *et al.*, 1996; Janko *et al.*, 1997; Chung *et al.*, 2000). In neural nets, the training procedure is less well defined and sources of information used by the net are less clear than those in the approach described here.

G. Comparison to other approaches for sound localization

The general approach employed here is not restricted to the specific ILD and IPD analysis carried out here, but is also applicable to more specific computational models of human binaural signal processing. Small nonlinearities in the extraction of binaural information by the models is acceptable for this type of analysis, as it has been shown here that the physically defined parameters are nonlinear functions of the sound field as well. Additional nonlinearities induced by the models (e.g., level dependencies) could add some additional uncertainty, which is processed by the fuzzy information processing strategy proposed here in the same way as the nonlinearity in the physical parameters.

V. CONCLUSIONS

In noise conditions, the observed random variation of short-term, narrow-band interaural parameters (ILDs and IPDs) with time is large compared to the systematic variation induced by a change of direction of the sound source of several degrees in azimuth. Additionally, noise fields cause a systematic shift of the average values of these parameters. Because of the stochastic temporal variability, integration of information across frequency, or time, or both, is necessary to estimate directions from interaural parameters in such conditions.

A way to achieve this integration is the combination of statistical information across frequency. A Bayesian approach was used for this, which takes the estimated probability density functions (PDFs) of ILDs and IPDs from a reference noise condition as *a priori* information. These *a priori* PDFs were measured and evaluated for a large number of conditions. The shapes of the observed distributions depend mainly on the SNR, azimuth, and elevation. The noise environment has a smaller influence on the shape of the distributions. This influence is most notable at medium SNRs. Using the Bayesian approach, the azimuth and elevation can be estimated robustly. The elevation can be estimated even in the median plane at SNRs as low as 5 dB. The localization performance depends mainly on the SNR in the test condition.

The high level of external noise in combination with the hypothesis of integrating probabilities in the neural system could explain why tuning curves of neurons sensitive to interaural timing found in physiological measurements are broad, unspecific, and often asymmetric, while the behavioral localization performance is robust and accurate. External noise with realistic statistical properties should be explicitly considered in physiological measurements and models of binaural processing.

ACKNOWLEDGMENTS

We are grateful to Birger Kollmeier for his substantial support and contribution to this work. We thank the members of the Oldenburg Medical Physics Group, especially Thomas Wittkop, Stephan Albani, and Jörn Otten, for providing tech-

nical support and for important discussions. Ronny Meyer prepared additional material which helped to discuss the results. Also, we are grateful to the staff of the Hearing Research Center at the Department of Biomedical Engineering, Boston University, for fruitful and motivating discussions during a visit from the second author. Thanks to Armin Kohlrausch, Fred Wightman, two anonymous reviewers, Steve Greenberg, Hermann Wagner, and Jesko Verhey for helpful suggestions and comments on earlier versions of this manuscript. This work was supported by DFG (European Graduate School Psychoacoustics), BMBF (Center of Excellence on Hearing Technology, 01 EZ 02 12), and DFG (SFB TR 31).

APPENDIX: MOMENT COEFFICIENTS AND PARAMETERS OF DISTRIBUTIONS OF CYCLIC RANDOM VARIABLES

The first moment μ_1 of the PDF $f(\theta)$ of a circular variable θ is

$$\mu_1 = \int_{-\pi}^{\pi} e^{i\theta} f(\theta) d\theta \tag{A1}$$

$$= \varrho e^{i\phi}. \tag{A2}$$

The argument ϕ of μ_1 denotes the *mean phase angle* and the absolute value ϱ denotes the *vector strength* or *resultant length*. The *circular standard deviation* σ_z is defined as

$$\sigma_z = \sqrt{-2 \log \varrho} \tag{A3}$$

(Fisher, 1993). The *circular variance* ν is defined as $\nu = 1 - \varrho$.

The vector strength can assume values between 0 and 1. If ϱ equals 1, the distribution has the shape of a delta function and all phase values are coincident. By way of contrast, $\varrho = 0$ could mean that the random variable is uniformly distributed at all phase values, or that the distribution has two peaks at an angular difference of π, for example.

Analogous to the central moments for the linear case, trigonometric central moments μ_p of order p can be defined as

$$\mu_p = \int_{-\pi}^{\pi} e^{ip(\theta-\phi)} f(\theta) d\theta. \tag{A4}$$

The imaginary part of the second central trigonometric moment $\mathcal{T}[\mu_2]$ can be used to calculate the *circular skew*

$$s_z = \frac{\mathcal{T}[\mu_2]}{\nu^{3/2}}, \tag{A5}$$

and the real part $\mathcal{R}[\mu_2]$ defines the *circular kurtosis*:

$$K_z = \frac{\mathcal{R}[\mu_2] - \varrho^4}{\nu^2}. \tag{A6}$$

Using these quantities it is possible to describe distributions of circular variables by a few descriptive parameters as in the linear case.

[1]Taking into account the effect of the 400-point Hann window, an effective average bandwidth of 0.96 ERB results.

[2]Probabilities $P(\Delta)$ are given in capital letters here and can be calculated by multiplying the probability density $p(\Delta)$ with the parameter interval $\delta\Delta$. This factor is omitted here, because it is constant and does not change the results.

[3]Specifically, the reference distributions were clustered using the hierarchical Ward technique (Ward, 1963; Kopp, 1978) so that for each of the 36 980 histograms (43 frequencies, 430 directions, and 2 parameters), 1 out of 550 samples of the marginal distributions was used. For simplicity, the influence of this data reduction technique on the localization accuracy is not discussed here. However, its application shows that a noticeable reduction of the *a priori* information is possible.

Albani, S., Peissig, J., and Kollmeier, B. (**1996**). "Model of binaural localization resolving multiple sources and spatial ambiguities," in *Psychoacoustics, Speech and Hearing Aids*, edited by B. Kollmeier (World Scientific Publishing, Singapore), pp. 227–232.

Allen, J. B. (**1977**). "Short term spectral analysis, synthesis and modification by discrete Fourier transform," IEEE Trans. Acoust., Speech, Signal Process. **ASSP-25**(3), 235–238.

Allen, J. B., and Rabiner, L. R. (**1977**). "A unified approach to short-time Fourier analysis and synthesis," in *Proceedings of the IEEE*, Vol. **65** (IEEE, New York).

Blauert, J. (**1983**). *Spatial Hearing—The Psychophysics of Human Sound Localization* (MIT, Cambridge, MA).

Bodden, M. (**1993**). "Modeling human sound source localization and the cocktail-party-effect," Acta Acust. (Beijing) **1**(1), 43–55.

Bodden, M. (**1996**). "Auditory demonstrations of a cocktail-party-processor," Acta Acust. (Beijing) **82**(2), 356–357.

Borisyuk, A., Semple, M. N., and Rinzel, J. (**2002**). "Adaptation and inhibition underlie responses to time-varying interaural phase cues in a model of inferior colliculus neurons," J. Neurophysiol. **88**(4), 2134–2146.

Braasch, J. (**2002a**). "Auditory Localization and Detection in Multiple-Sound Source Scenarios," Ph.D. thesis, Ruhr-Universität Bochum, Düsseldorf, VDI-Verlag. Fortschritts-Berichte VDI Reihe 10 Nr. 707.

Braasch, J. (**2002b**). "Localization in presence of a distracter and reverberation in the frontal horizontal plane. II. Model algorithms," Acta. Acust. Acust. **88**, 956–969.

Braasch, J., and Hartung, K. (**2002**). "Localization in presence of a distracter and reverberation in the frontal horizontal plane. I. Psychoacoustical data," Acta. Acust. Acust. **88**, 942–955.

Brainard, M. S., Knudsen, E. I., and Esterly, S. D. (**1992**). "Neural derivation of sound source location: Resolution of spatial ambiguities on binaural cues," J. Acoust. Soc. Am. **91**, 1015–1027.

Breebaart, J., van de Par, S., and Kohlrausch, A. (**1999**). "The contribution of static and dynamically varying ITDs and IIDs to binaural detection," J. Acoust. Soc. Am. **106**, 979–992.

Brugge, J. F. (**1992**). "An overview of central auditory processing," in *The Mammalian Auditory Pathway: Neurophysiology*, edited by A. N. Popper and R. R. Fay, Vol. **2** of *Springer Handbook on Auditory Research* (Springer Verlag, New York), Chap. 1, pp. 1–33.

Butler, R. A. (**1986**). "The bandwidth effect on monaural and binaural localization," Hear. Res. **21**(1), 67–73.

Caird, D., and Klinke, R. (**1987**). "Processing of interaural time and intensity differences in the cat inferior colliculus," Exp. Brain Res. **68**(2), 379–392.

Chung, W., Carlile, S., and Leong, P. (**2000**). "A performance adequate computational model for auditory localization," J. Acoust. Soc. Am. **107**, 432–445.

Clarey, J. C., Barone, P., and Imig, T. J. (**1992**). "Physiology of thalamus and cortex," in *The Mammalian Auditory Pathway: Neurophysiology*, edited by A. N. Popper and R. R. Fay, Vol. **2** of *Springer Handbook on Auditory Research* (Springer Verlag, New York), Chap. 5, pp. 232–334.

Colburn, H. S. (**1996**). "Computational models of binaural processing," in *Auditory Computation*, edited by H. L. Hawkins, T. A. McMullen, A. N. Popper, and R. R. Fay, Vol. **6** of *Springer Handbook of Auditory Research* (Springer, New York), Chap. 8, pp. 332–400.

Datum, M. S., Palmieri, F., and Moiseff, A. (**1996**). "An artificial neural-network for sound localization using binaural cues," J. Acoust. Soc. Am. **100**, 372–383.

Domnitz, R. H., and Colburn, H. S. (**1976**). "Analysis of binaural detection models for dependence on interaural target parameters," J. Acoust. Soc. Am. **59**, 598–601.

Duda, R. O. (**1997**). "Elevation dependence of the interaural transfer function," in *Binaural and Spatial Hearing in Real and Virtual Environments*, edited by R. H. Gilkey and T. R. Anderson (Erlbaum, Mahwah, NJ), Chap. 3, pp. 49–75.

Durlach, N. I., and Colburn, H. S. (**1978**). "Binaural phenomena," in *Handbook of Perception—Hearing*, edited by E. C. Carterette and M. P. Friedman (Academic, New York), Vol. **4**, Chap. 10, pp. 365–466.

Fisher, N. I. (**1993**). *Statistical Analysis of Circular Data* (Cambridge U.P., Cambridge).

Fitzpatrick, D., Batra, R., and Stanford, T. (**1997**). "A neural population code for sound localization," Nature (London) **388**, 871–874.

Good, M. D., and Gilkey, R. H. (**1996**). "Sound localization in noise: The effect of signal-to-noise ratio," J. Acoust. Soc. Am. **99**, 1108–1117.

Hancock, K. E., and Delgutte, B. (**2004**). "A physiologically based model of interaural time difference discrimination," J. Neurosci. **24**(32), 7110–7117.

Harper, N. S., and McAlpine, D. (**2004**). "Optimal neural population coding of an auditory spatial cue," Nature (London) **430**, 682–686.

Henning, G. B. (**1973**). "Effect of interaural phase on frequency and amplitude discrimination," J. Acoust. Soc. Am. **54**, 1160–1178.

Ito, Y., Colburn, H. S., and Thompson, C. L. (**1982**). "Masked discrimination of interaural time delays with narrow-band signal," J. Acoust. Soc. Am. **72**, 1821–1826.

Janko, J. A., Anderson, T. R., and Gilkey, R. H. (**1997**). "Using neural networks to evaluate the viability of monaural and interaural cues for sound localization," in *Binaural and Spatial Hearing in Real and Virtual Environments*, edited by R. H. Gilkey and T. R. Anderson (Erlbaum, Mahwah, NJ), Chap. 26, pp. 557–570.

Jeffress, L. A. (**1948**). "A place theory of sound localization," J. Comp. Physiol. Psychol. **41**, 35–39.

Jenison, R. L. (**2000**). "Correlated cortical populations can enhance sound localization performance," J. Acoust. Soc. Am. **107**, 414–421.

Joris, P. X., and Yin, T. C. T. (**1996**). "Envelope coding in the lateral superior olive. I. Sensitivity to interaural time difference," J. Neurophysiol. **73**(3), 1043–1062.

Knudsen, E. I. (**1982**). "Auditory and visual maps of space in the optic tectum of the owl," J. Neurosci. **2**, 1177–1194.

Kollmeier, B., and Koch, R. (**1994**). "Speech enhancement based on physiological and psychoacoustical models of modulation perception and binaural interaction," J. Acoust. Soc. Am. **95**, 1593–1602.

Kollmeier, B., Peissig, J., and Hohmann, V. (**1993**). "Real-time multi-band dynamic compression and noise reduction for binaural hearing aids," J. Rehabil. Res. Dev. **30**(1), 82–94.

Kopp, B. (**1978**). "Hierarchical classification III: Average-linkage, median, centroid, WARD, flexible strategy," Biom. J. **20**(7/8), 703–711.

Kuwada, S., and Yin, T. C. T. (**1983**). "Binaural interaction in low-frequency neurons in inferior colliculus of the cat. I. Effects of long interaural delays, intensity, and repetition rate on interaural delay function," J. Neurophysiol. **50**(4), 981–999.

Kuwada, S., and Yin, T. C. T. (**1987**). "Physiological studies of directional hearing," in *Directional Hearing*, edited by W. A. Yost and G. Gourevitch (Springer, New York), Chap. 6, pp. 146–176.

Liu, C., Wheeler, B. C., O'Brien, Jr., W. D., and Bilger, R. C. (**2000**). "Localization of multiple sound sources with two microphones," J. Acoust. Soc. Am. **108**, 1888–1905.

Lyon, R. F. (**1983**). "A computational model of binaural localization and separation," in *Proceedings of the International Conference on Acoustics, Speech and Signal Processing ICASSP'83*, Vol. **3** (IEEE, New York).

Macpherson, E. A., and Middlebrooks, J. C. (**2002**). "Listener weighting of cues for lateral angle: The duplex theory of sound localization revisited," J. Acoust. Soc. Am. **111**, 2219–2236.

Malone, B. J., Scott, B. H., and Semple, M. N. (**2002**). "Context-dependent adaptive coding of interaural phase disparity in the auditory cortex of awake macaques," J. Neurosci. **22**(11), 4625–4638.

McAlpine, D., and Grothe, B. (**2003**). "Sound localization and delay lines—do mammals fit the model?" Trends Neurosci. **26**(7), 347–350.

Moore, B. C. J. (**1989**). *An Introduction to the Psychology of Hearing*, 3rd ed. (Academic, New York), Vol. **1**, Chap. 3, pp. 100–101.

Nakashima, H., Chisaki, Y., Usagawa, T., and Ebata, M. (**2003**). "Frequency domain binaural model based on interaural phase and level differences," Acoust. Sci. & Tech. **24**(4), 172–178.

Neti, C., Young, E. D., and Schneider, M. H. (**1992**). "Neural network models of sound localization based on directional filtering by the pinna," J. Acoust. Soc. Am. **92**, 3140–3156.

Otten, J. (2001). "Factors influencing acoustical localization," Ph.D. thesis, Universität Oldenburg, Oldenburg, Germany.

Sachs, L. (**1992**). *Angewandte Statistik (Applied Statistics)*, 7 ed. (Springer, Berlin).

Schauer, C., Zahn, T., Paschke, P., and Gross, H. M. (**2000**). "Binaural sound localization in an artificial neural network," in *Proc. International Conference on Acoustics, Speech and Signal Processing (ICASSP) 2000*, Vol. **2** (IEEE, New York).

Searle, C. L., Braida, L. D., Cuddy, D. R., and Davis, M. F. (**1975**). "Binaural pinna disparity: Another localization cue," J. Acoust. Soc. Am. **57**, 448–455.

Shackleton, T., Meddis, R., and Hewitt, M. J. (**1992**). "Across frequency integration in a model of lateralization," J. Acoust. Soc. Am. **91**, 2276–2279.

Shaw, E. A. (**1997**). "Acoustical features of the human external ear," in *Binaural and Spatial Hearing in Real and Virtual Environments*, edited by R. H. Gilkey and T. R. Anderson (Erlbaum, Mahwah, NJ), Chap. 2, pp. 25–47.

Spitzer, M. W., and Semple, M. N. (**1991**). "Interaural phase coding in auditory midbrain: Influence of dynamic stimulus features," Science **254**(5032), 721–724.

Stern, R. M., and Bachorski, R. J. (**1983**). "Dynamic cues in binaural perception," in *Hearing—Physiological Bases and Psychophysics*, edited by R. Klinke and R. Hartmann (Springer, Heidelberg), pp. 209–215.

Stern, R. M., and Trahiotis, C. (**1997**). "Models of binaural perception," in *Binaural and Spatial Hearing in Real and Virtual Environments*, edited by R. H. Gilkey and T. R. Anderson (Erlbaum, Mahwah, NJ), Chap. 24, pp. 499–531.

Stern, R. M., Zeiberg, A. S., and Trahiotis, C. (**1988**). "Lateralization of complex binaural stimuli: A weighted-image model," J. Acoust. Soc. Am. **84**, 156–165.

Stern, Jr., R. M., Slocum, J. E., and Phillips, M. S. (**1983**). "Interaural time and amplitude discrimination in noise," J. Acoust. Soc. Am. **73**, 1714–1722.

Tribolet, J. M. (**1977**). "A new phase unwrapping algorithm," IEEE Trans. Acoust., Speech, Signal Process. **ASSP-25**(2), 170–177.

Wagner, H. (**1991**). "A temporal window for lateralization of interaural time difference by barn owls," J. Comp. Physiol., A **169**, 281–289.

Ward, Jr., J. H. (**1963**). "Hierarchical grouping to optimize an objectice function," J. Am. Stat. Assoc. **58**(301), 236–244.

Wightman, F. L., and Kistler, D. J. (**1989a**). "Headphone simulation of free-field listening. I: Stimulus synthesis," J. Acoust. Soc. Am. **85**, 858–867.

Wightman, F. L., and Kistler, D. J. (**1989b**). "Headphone simulation of free-field listening. II: Psychophysical validation," J. Acoust. Soc. Am. **85**, 868–878.

Wittkop, T., Albani, S., Hohmann, V., Peissig, J., Woods, W. S., and Kollmeier, B. (**1997**). "Speech processing for hearing aids: Noise reduction motivated by models of binaural interaction," Acust. Acta Acust. 83(4), 684–699.

Zurek, P. M. (**1991**). "Probability distributions of interaural phase and level differences in binaural detection stimuli," J. Acoust. Soc. Am. **90**, 1927–1932.