PROCEEDINGS of the
23rd International Congress on Acoustics

9 to 13 September 2019 in Aachen, Germany

A localization algorithm based on head-related transfer functions

Maike GERHARD(1), Patrick SCHILLBERG(2), Hermann WAGNER(2), Hartmut FÜHR(3)

(1)Lehrstuhl A für Mathematik, RWTH Aachen University, Germany, maike.gerhard@matha.rwth-aachen.de
(2)Institute of Biology, RWTH Aachen University, Germany
(1)Lehrstuhl A für Mathematik, RWTH Aachen University, Germany

Abstract
A localization algorithm based on head-related transfer functions (HRFTs) is introduced. The algorithm minimizes the Euclidean distance to direction-dependent binaural signal spaces, in order to find an estimator of the actual sound source position. Numerical experiments with different target sounds (clicks, white noise, speech, rustling of leaves) and two signal-to-noise ratios (10 dB and 0 dB), where the masker is non-spatialized white noise, show that this projection algorithm outperforms template matching and the cross-channel algorithm in nearly all experimental conditions. In particular, the algorithm is robust to changes of the emitted signal's phase spectrum, unlike template matching.
For white noise maskers, it is possible to compute SNR dependent estimates for the error probability in the task of discriminating two directions, based on the associated HRTFs. We present simulations that demonstrate the precision of the estimates. We show how these probabilities can be employed as a means to mathematically analyze HRTFs, in particular with the aim of predicting localization performance from the HRTF data set.
Keywords: Sound localization, HRTFs

1 INTRODUCTION

For the localization of sound source positions, humans and animals use various spatial cues contained in the two sound signals reaching the left and right ear. It is currently assumed that the interaural time difference (ITD), the interaural level difference (ILD), and high frequency spectral notches are most important among these cues [1, 2]. All these cues and perhaps even more spatial information can be extracted from *head-related impulse responses* (HRIRs) [3], which encode the linear direction dependent acoustic filtering of sounds by the listener's outer ears, head, and body. This acoustic filtering models physical phenomena such as interference, reflection, and diffraction and depends on both the direction of arrival and the listener's anthropometric parameters. The frequency domain representation of a HRIR is the head-related transfer function (HRTF), obtained as the Fourier transform of the HRIR.

In the past decades, HRTFs have been investigated from a variety of perspectives. A topic related to all directions of research is the mathematical analysis of HRTF data. The present paper is a further contribution to this research by deepening the theoretical understanding of the acoustic information inherent in head-related transfer functions. Starting from first principles, a sound localization algorithm based on head-related transfer functions is developed. Numerical experiments with different target sounds show, that this algorithm performs highly efficiently, also in comparison to other algorithms. We also introduce a theoretical estimate for the error probability in the task of discriminating two directions. This estimate depends on the signal-to-noise ratio (SNR) of the emitted signal as well as on the HRTFs associated to the two positions taking part in the discrimination task.

Section 2 introduces a new HRTF based sound localization algorithm. At the beginning of this section we describe the model of acoustic signal transmission that forms the basis of the considerations throughout this paper. We also introduce the mathematical framework by giving the necessary definitions. Afterwards we present a mathematical development of the mentioned localization algorithm. The algorithm performs orthogonal projections of the incoming binaural sound onto HRTF-dependent binaural subspaces and minimizes the Euclidean distance between these projections and the original sound.

Section 3 compares the projection algorithm to two other algorithms: The template matching algorithm and the cross-channel algorithm introduced by MacDonald [5]. First we give an overview of the operating modes of the two competing algorithms. Then we compare the performances of the three algorithms by conducting numerical experiments implemented in MATLAB. Two different signal-to-noise ratios (10 dB and 0 dB) and several target sounds were used in these experiments to simulate various localization scenarios.
In section 4 we consider 2-choice discrimination tasks, where the listener has to decide which of two given positions is the actual sound source position. If white noise maskers are used in such paradigms, it is possible to compute an upper bound for the error probability. This bound depends on the HRTFs associated to the two positions as well as on the SNR of the incoming sound. We show, that this bound can be used to predict the localization performance of individuals.
The paper's conclusion in section 5 sums up the main results of our work.

2 A NOVEL HRTF-BASED LOCALIZATION ALGORITHM

In this section we introduce a new algorithm for sound localization, called projection algorithm. Prior to that, we present the mathematical framework.

2.1 Model assumptions and mathematical basics

The basic problem of sound localization can be formulated as follows: An acoustic signal x is emitted from direction P. The binaural sound (x_l, x_r) arriving at the left ear and the right ear is then described by the equations

$$x_l = x * H_{l,P} + \varepsilon_l \ , \ x_r = x * H_{r,P} + \varepsilon_r \ .$$

Here

- $H_{l,P}$ and $H_{r,P}$ denote the left and right head-related impulse responses associated to position P,
- the symbol $*$ denotes convolution, and
- $\varepsilon_l, \varepsilon_r$ is interfering noise that is not related to the emitted signal x.

The task of sound localization consists in estimating P from the pair (x_l, x_r), usually without any prior knowledge of the emitted signal, or of the interfering noise.
When we study sound localization as a discrimination task, we ask whether two directions P and Q can be sufficiently discriminated, for example in the setting of a behavioral experiment. Since our analysis is purely based on the available physical *acoustic* information contained in the HRIRs, we do not take into account any prior knowledge of the emitted signal or the masker. Thus, in view of the above equation, the pertinent question becomes whether

$$x * H_{l,P} \approx y * H_{l,Q} \qquad \text{and} \qquad x * H_{r,P} \approx y * H_{r,Q}$$

hold with a suitably chosen signal y emitted from Q. If this is the case, it becomes very difficult to decide whether the signal arrives from P or Q.

For the mathematical modeling we assume that all acoustic objects (signals, HRIRs, noise, etc.) have time domain representations of the same length $N \in \mathbb{N}$ and are measured with the same sampling rate $S = \frac{1}{\Delta_t}$. Hence the space of acoustic signals is given by $l^2(\underline{N}) \cong \mathbb{C}^N$ with $\underline{N} := \{0, 1, \ldots, N-1\}$. The kth entry of each vector corresponds to the time $k \cdot \Delta_t$. The energy of a signal x is given by its squared Euclidean norm $\|x\|^2 = \sum_{k=0}^{N-1} |x(k)|^2$ and the (cyclic) convolution of two such signals is given by

$$(x * y)(k) = \sum_{n=0}^{N-1} x(n) y((k-n) \mod N), \quad k \in \underline{N}.$$

The Fourier transform $\mathscr{F}$ converts signals from time domain to frequency domain:

$$\mathscr{F}: l^2(\underline{N}) \to l^2(\underline{N}), \quad \mathscr{F}(x)(k) = \hat{x}(k) = \sum_{n=0}^{N-1} x(n) e^{-2\pi ikn/N}, \quad k \in \underline{N}.$$

Hence all frequency domain representations also have length N. Convolution in time domain corresponds to multiplication in frequency domain and $\mathscr{F}$ conserves energy, at least up to a fixed factor:

$$\widehat{(x*y)} = \hat{x}\hat{y} \quad \text{and} \quad N\|x\|^2 = \|\hat{x}\|^2.$$

Here and in the following multiplication and division is componentwise.

2.2 The projection algorithm

As mentioned before the task of sound localization consists in estimating the actual sound source position P_0 from the binaural sound

$$(x_l, x_r) = (x * H_{l,P_0} + \varepsilon_l,\ x * H_{r,P_0} + \varepsilon_r) \in l^2(\underline{N}) \times l^2(\underline{N}).$$

Since we assume that the emitted signal x and the noise $(\varepsilon_l, \varepsilon_r)$ are unknown, the HRTFs have to be used for solving this task. We utilize the fact, that in the case of a noise-free signal transmission all signals arriving from the very same position have the same pattern:

Definition 2.1. *Given a position P, the space composed of all binaural sounds arriving from that position P is denoted by $\mathscr{H}_\text{P}$, i.e.*

$$\mathscr{H}_\text{P} = \left\{ (x * H_{l,P}, x * H_{r,P}) \mid x \in l^2(\underline{N}) \right\} = \left\{ (\hat{x}\hat{H}_{l,P}, \hat{x}\hat{H}_{r,P}) \mid \hat{x} \in l^2(\underline{N}) \right\}.$$

For every position P the set $\mathscr{H}_\text{P}$ is a subspace of the space of all binaural sounds $l^2(\underline{N}) \times l^2(\underline{N})$. In addition, for a second position Q different from P one expects $\mathscr{H}_\text{Q} \neq \mathscr{H}_\text{P}$, since the HRTFs associated to the two positions will not be equal, or in more precisely: the quotients $\hat{H}_{l,P}/\hat{H}_{r,P}$ and $\hat{H}_{l,Q}/\hat{H}_{r,Q}$ will not be equal. That means in the noise-free case, the HRTFs describe subspaces of the space of spatialized sounds. For correct localization of an incoming sound it is then sufficient to determine which of these subspaces contains the sound. Unfortunately the noisy incoming signal cannot be expect to lie in any of the subspaces $\mathscr{H}_\text{P}$. However, we can compute the Euclidean distances between the incoming sound and the different subspaces $\mathscr{H}_\text{P}$, and minimize of all available positions P to obtain an estimator of the actual sound source position:

Algorithm 2.2 (Projection algorithm)**.** ***Input:*** *A binaural sound $(x_l, x_r) \in l^2(\underline{N}) \times l^2(\underline{N})$.*
Output: *The estimator P_{est} of the sound source position P_0 given by*

$$P_{est} = \operatorname{argmin}_P \operatorname{dist}\left((x_l, x_r), \mathscr{H}_P\right).$$

The theorem below shows, that the distance $\operatorname{dist}((x_l, x_r), \mathscr{H}_P)$ is computable in an explicit and efficient way. The overall minimization task is solved subsequently by running a simple loop through all available directions P.

Theorem 2.3. *For a position P the quantity* $\operatorname{dist}((x_l, x_r), \mathscr{H}_P)$ *can be computed explicitly.*

Proof. We have to solve

$$\min_{x_P \in l^2(\underline{N})} \left\| (x_l, x_r) - (x_P * \hat{H}_{l,P}, x_P * \hat{H}_{r,P}) \right\|^2.$$

Because of the energy conservation of the Fourier transform, we can solve this in the frequency domain, where it becomes

$$\min_{\hat{x}_P \in l^2(\underline{N})} \left\| (\hat{x}_l, \hat{x}_r) - (\hat{x}_P \hat{H}_{l,P}, \hat{x}_P \hat{H}_{r,P}) \right\|^2 = \min_{\hat{x}_P \in l^2(\underline{N})} \sum_{k \in \underline{N}} \left\| (\hat{x}_l(k), \hat{x}_r(k)) - \hat{x}_P(k)(\hat{H}_{l,P}(k), \hat{H}_{r,P}(k)) \right\|^2.$$

Now the minimization can be done for every term of the sum (i.e. for every frequency) individually. For every $k \in \underline{N}$ we project the vector $(\hat{x}_l(k), \hat{x}_r(k)) \in \mathbb{C}^2$ orthogonal onto the closed subspace

$$\{\hat{x} \cdot (\hat{H}_{l,P}(k), \hat{H}_{r,P}(k)) \mid \hat{x} \in \mathbb{C}\} \subset \mathbb{C}^2.$$

The solution is given by

$$\hat{x}_P(k) = \frac{(\hat{x}_l(k)\overline{\hat{H}_{l,P}(k)} + \hat{x}_r(k)\overline{\hat{H}_{r,P}(k)})}{|\hat{H}_{l,P}(k)|^2 + |\hat{H}_{r,P}(k)|^2} \quad \text{for } k \in \underline{N},$$

where the overline denotes the complex conjugate. □

We note that in the noise-free case, directions P and Q can only be confused by the algorithm if there are signals $x_P \neq 0$ and $x_Q \neq 0$ such that x_P emitted from P coincides with x_Q emitted from Q. Mathematically, this can be expressed via the following equivalent conditions:

$$\mathscr{H}_{\mathrm{P}} \cap \mathscr{H}_{\mathrm{Q}} \supsetneq \{0\} \quad \Leftrightarrow \quad \exists\ k \in \underline{N}: \ \frac{\hat{H}_{l,P}(k)}{\hat{H}_{r,P}(k)} = \frac{\hat{H}_{l,Q}(k)}{\hat{H}_{r,Q}(k)}.$$

Note that the right hand side of the equivalence revals that typical signals leading to nonempty intersections are tones.

3 PERFORMANCE OF THE PROJECTION ALGORITHM

In this section we compare the performance of the projection algorithm to the performance of two other sound localization algorithms: Template matching and the cross-channel algorithm. Note that all of these algorithms require the complete HRIR data set of a listener in order to localize the sound source.

3.1 Template matching

If the emitted signal x is a click, the binaural sound (x_l, x_r) arriving at the ears is up to a fixed factor equal to the HRIRs (H_{l,P_0}, H_{r,P_0}) associated to the actual sound source position P_0. This observation leads to a simple localization technique called *template matching*, which determines which pair of HRIRs $(H_{l,P}, H_{r,P})$ is most similar to the incoming sound. All correlation-based localization algorithms perform some sort of template matching. If the HRIRs themselves are used as templates, this investigation can be conveniently performed in the frequency domain. In general the energy of different HRTF pairs is also different. To ensure that this does not influence the outcome, all HRTFs were normalized before comparison with the incoming sound. So the template related to position P is given by

$$(T_{l,P}, T_{r,P}) = \frac{(\hat{H}_{l,P}, \hat{H}_{r,P})}{\sqrt{\left\|\hat{H}_{l,P}\right\|^2 + \left\|\hat{H}_{r,P}\right\|^2}}.$$

The real part of a scalar product $\langle v, w \rangle$ is an often used measure for the similarity of vectors v and w, which is also employed by the template matching algorithm.

Algorithm 3.1 (Template matching)**.** ***Input:*** *A binaural sound* $(x_l, x_r) \in l^2(\underline{N}) \times l^2(\underline{N})$.
Output: *The estimator* P_{est} *of the sound source position* P_0 *given by*

$$P_{est} = \mathrm{argmax}_P\ \Re\left(\langle T_{l,P}, \hat{x}_l \rangle + \langle T_{r,P}, \hat{x}_r \rangle\right).$$

We are not aware of sources studying sound localization by directly using HRTFs as templates for sound localization. However, the localization algorithm presented in [4] can be interpreted as as a template matching algorithm applied to suitably preprocessed binaural signals.

3.2 The cross-channel algorithm

The cross-channel algorithm was developed by Justin A. MacDonald [5]. This algorithm is based on a simple yet elegant idea: If a signal x is emitted from position P_0 the noise-free sounds arriving at the ears are

$$x_l = x * H_{l,P_0} \quad \text{and} \quad x_r = x * H_{r,P_0},$$

respectively. If we now convolve the left sound x_l with the right ear HRIR associated to position P_0 and the right sound x_r with the left ear HRIR associated to P_0, we get equal results, because of the commutativity and associativity of convolution:

$$x_l * H_{r,P_0} = \left(x * H_{l,P_0}\right) * H_{r,P_0} = \left(x * H_{r,P_0}\right) * H_{l,P_0} = x_r * H_{l,P_0}.$$

For any other position $P \neq P_0$, and general signals x, this cannot be expected:

$$x_l * H_{r,P} = \left(x * H_{l,P_0}\right) * H_{r,P} \neq \left(x * H_{r,P_0}\right) * H_{l,P} = x_r * H_{l,P}.$$

In the presence of noise, the equation $x_l * H_{r,P_0} = x_r * H_{l,P_0}$ is no longer true. Nevertheless it can be expected that the correlation between $x_l * H_{r,P_0}$ and $x_r * H_{l,P_0}$ is larger than the correlation between $x_l * H_{r,P}$ and $x_r * H_{l,P}$.

Algorithm 3.2 (Cross-channel algorithm)**.** ***Input:*** *A binaural sound* $(x_l, x_r) \in l^2(\underline{N}) \times l^2(\underline{N})$.
Output: *The estimator* P_{est} *of the sound source position* P_0 *given by*

$$P_{est} = \text{argmax}_P \; r\left(x_l * H_{r,P}, x_r * H_{l,P}\right),$$

where r denotes the Pearson correlation coefficient.

3.3 Comparison

We used a numerical localization experiment implemented in MATLAB (The MathWorks, Natick, MA) to compare the performance of the three algorithms. The spatial binaural sounds (x_l, x_r) used as inputs were generated by convolving a signal x with the HRIRs H_{l,P_0} and H_{r,P_0} associated to the simulated sound source position P_0. Afterwards non-spatialized white noise was added to the convolution product, to achieve the desired SNR. We used publically available HRIRs from the CIPIC database [6]. This database includes 1250 positions whose azimuth and elevation angles are measured in a head-centered interaural poplar coordinate system. All positions are located on a sphere with radius 1 m around the listeners head. The inputs (x_l, x_r) differed from each other with respect to

- the source position P_0: 35 sound source positions ranging in azimuth from -80 deg to 80 deg in steps of 40 deg (5 values for the azimuth) and in elevation from -45 deg to 225 deg in steps of 45 deg (7 values for the elevation) take part in the experiment.
- the SNR: 10 dB or 0 dB.
- the emitted signal x: We used 8 different signals, each with a length of 100 ms.
  1.) Signal with a phase spectrum equal to 0 and a flat amplitude spectrum over the frequency range $\Omega_{low} = [0 \text{ Hz}, \; 4000 \text{ Hz}]$.
  2.) The same as in 1.), but with a different frequency range $\Omega_{high} = [4000 \text{ Hz}, \; 22000 \text{ Hz}]$.
  3.) The same as in 1.), but with a different frequency range $\Omega_{broad} = [0 \text{ Hz}, \; 22000 \text{ Hz}]$.
  4.) The same as in 1.), but with a random phase spectrum.
  5.) The same as in 2.), but with a random phase spectrum.
  6.) The same as in 3.), but with a random phase spectrum.

7.) Speech (female voice).

8.) Rustling of leaves.

Each signal was combined with both SNRs and each signal/SNR-combination was emitted 50 times per position. The noise was random in each trial, but all algorithms were tested with exactly the same inputs.

As error measure we use the spherical error, which is the absolute value of the angular distance between the actual sound source position P_0 and the estimated position P_{est} in degrees. We performed the experiment described above with the HRIR sets of six different subjects. The mean spherical errors presented in table 1 are averaged across subjects and source positions.

Table 1. Mean spherical errors for inputs with a SNR of 10 dB (left) and 0 dB (right)

	10 dB			0 dB		
Signal	Projection	Template Matching	Cross-channel	Projection	Template Matching	Cross-channel
1.)	0.02	40.56	7.96	3.19	40.69	26.35
2.)	0.02	1.58	2.01	0.51	1.68	11.43
3.)	0.02	0.00	0.32	0.46	0.00	4.62
4.)	0.05	73.15	15.84	8.27	77.31	35.19
5.)	0.01	67.25	1.53	1.45	68.50	13.39
6.)	0.01	66.75	0.31	0.51	67.32	5.42
7.)	1.08	59.71	49.26	18.60	74.95	58.43
8.)	0.01	54.23	20.45	5.34	59.50	45.15

The table shows, that the projection algorithm outperforms the other algorithms in nearly all conditions. The only exception is signal 3.), which is best estimated by template matching. This is not surprising, since signal 3.) is a click and template matching is designed for such signals. A comparison of signals 1-3.) to signals 4-6.) emphasizes, that the success of template matching highly depends on the phase spectrum of the emitted signal, which is not the case for the projection algorithm. Note, that even at a low SNR of 0 dB the mean errors of the projection algorithm are with one exception (speech) smaller than 10 degrees.

4 ERROR PROBABILITY IN DISCRIMINATION TASKS

In this section we consider 2-choice discrimination tasks, where the listener has to decide whether the signal arrives from the actual sound source position P or from a second position, which we call Q. Assuming that the masker is white noise, we can use probability theory to compute an upper bound for the error probability in such tasks. This bound depends on the positions' HRTFs and the sound's SNR.

In this context we need a stochastic rather than a deterministic model: In the frequency domain the binaural sound $(\hat{x}_l, \hat{x}_r) \in l^2(\underline{N}) \times l^2(\underline{N})$ is given by

$$\hat{x}_l = \hat{H}_{l,P} \cdot \hat{x} + \sigma\ \hat{\varepsilon}_l \quad \text{and} \quad \hat{x}_r = \hat{H}_{r,P} \cdot \hat{x} + \sigma\ \hat{\varepsilon}_r,$$

where $\sigma \in \mathbb{R}_{\geq 0}$ and $\hat{\varepsilon}_l = (\hat{\varepsilon}_{l,0}, \ldots, \hat{\varepsilon}_{l,N-1})$, $\hat{\varepsilon}_r = (\hat{\varepsilon}_{r,0}, \ldots, \hat{\varepsilon}_{r,N-1})$ are random vectors with

$$\Re(\hat{\varepsilon}_{l,i}), \Im(\hat{\varepsilon}_{l,i}), \Re(\hat{\varepsilon}_{r,i}), \Im(\hat{\varepsilon}_{r,i}) \quad \overset{iid}{\sim} \quad \mathscr{N}(0, \alpha^2) \quad \text{for all } i \in \underline{N}.$$

This means that the real and imaginary parts of the scalar components of $\hat{\varepsilon}_l$ and $\hat{\varepsilon}_r$ are independent and identically normal distributed with mean 0 and variance α^2. We make two assumptions:

$$\|(\hat{H}_{l,P}\hat{x}, \hat{H}_{r,P}\hat{x})\|^2 = 1 \quad \text{and} \quad \mathbb{E}(\|(\hat{\varepsilon}_l, \hat{\varepsilon}_r)\|^2) = 1,$$

where $\mathbb{E}$ denotes the expected value. Because of the second assumption the SNR of the binaural sound is determined by the scalar σ. Let $\mathbb{P}(Q|P)$ denote the error probability, i.e.

$$\mathbb{P}(Q|P) \mathrel{\hat{=}} \text{the probability that the listener chooses } Q \text{ althoug the signal arrives from } P.$$

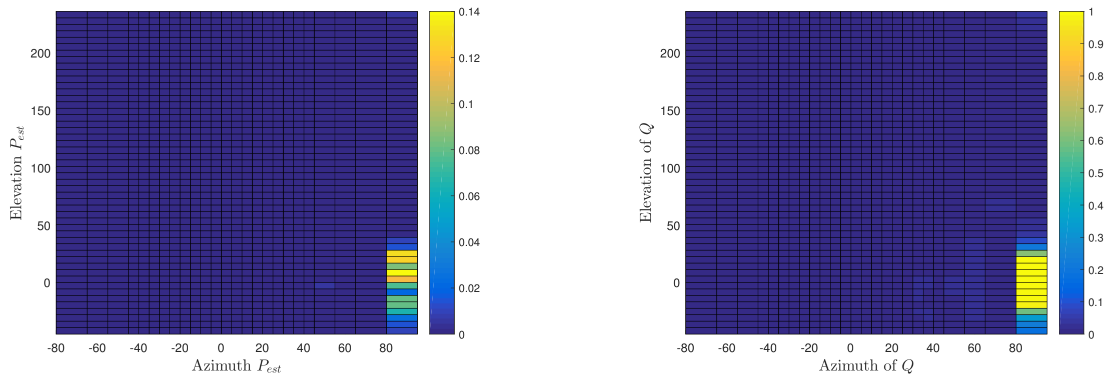
Then the following theorem holds true.

Theorem 4.1. *Let $\{u_1, \ldots, u_N\}$ be an orthonormal basis of $\mathscr{H}_Q^{\perp}$ and $\{v_1, \ldots, v_N\}$ an orthonormal basis of $\mathscr{H}_P^{\perp}$. Then we have*

$$\mathbb{P}(Q|P) \leq \mathrm{UB}(P,Q) := \min\left\{1,\ \left(\frac{1}{2N} - \frac{1}{2N^2}\sum_{i=1}^{N}\sum_{j=1}^{N}|\langle u_i, v_j\rangle|^2\right)\left(\frac{\sigma}{\|z\|}\right)^4 + \frac{1}{N}\left(\frac{\sigma}{\|z\|}\right)^2\right\}$$

with $z := \Pi_{\mathscr{H}_Q^{\perp}}\left((\hat{H}_{l,P}\hat{x}, \hat{H}_{r,P}\hat{x})\right)$ and $\Pi_{\mathscr{H}_Q^{\perp}}$ denotes the orthogonal projection onto $\mathscr{H}_Q^{\perp}$.

For $P = Q$ we define $\mathrm{UB}(P,Q) = 1$. To examine whether the quantity $\mathrm{UB}(P,Q)$ can be used to predict the results of localization experiments, we compare $\mathrm{UB}(P,Q)$ to the performance of the projection algorithm. Two examples are shown in figures 1 and 2.



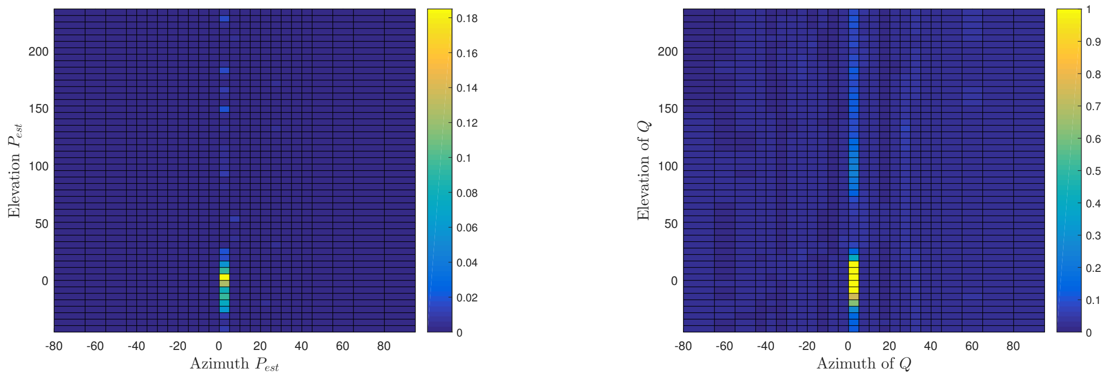
(a) Localization performance

(b) UB

Figure 1. Comparison between the results of a localization experiment (a) and $\mathrm{UB}(P_0, Q)$ (b). The sound source position is $P_0 = (80, 0)$, the emitted signals are of type 1.), and the SNR is -5 dB.

Part (a) of figure 1 shows the results of a localization experiment implemented in MATLAB. In this experiment the source position $P_0 = (80, 0)$ was fixed. The emitted signals were of type 1.) (see section 3.3). The masker was non-spatialized white noise and the SNR was -5 dB. We performed 200 localization runs with random noise, respectively. The color map displays the relative frequency with that a position, indicated by x- and y-axis, was estimated as sound source position by the projection algorithm. Part (b) of the same figure shows the values of $\mathrm{UB}(P_0, Q)$ for a fixed $P_0 = (80, 0)$. Azimuth and elevation of position Q are indicated by the x- and y-axis, respectively. The value of $\mathrm{UB}(P_0, Q)$ is indicated by the color. For the calculation of $\mathrm{UB}(P_0, Q)$ we used the same SNR (-5 dB) and signal (type 1.)) as in the localization experiment related to part (a), to ensure that the comparison is valid.

Figure 2 is of the same type as figure 1. Here the sound source position is given by $P_0 = (0, 0)$ and the signals are of type 3.). Furthermore the SNR is even smaller, namely -11 dB.



(a) Localization performance

(b) UB

Figure 2. Comparison between the results of a localization experiment (a) and $\mathrm{UB}(P_0, Q)$ (b). The sound source position is $P_0 = (0, 0)$, the emitted signals are of type 3.), and the SNR is -11 dB.

Clearly, both figures show pairs of very similar plots. This can also be confirmed quantitatively: The correlation coefficients between the maps in figure 1 and 2 are given by 0.92 and 0.91, respectively. Nevertheless, the algorithm performs a little better than predicted by UB. For example, the uncertainty in elevation shown in part (b) of figure 2 is less pronounced in part (a) of the same figure. That was to be expected, since UB is an upper bound for the error probability.

5 CONCLUSIONS

The main purpose of the present paper is the introduction of a new HRTF based sound localization algorithm. The presented results stress that this algorithm exhibits a strong performance in comparison to established algorithms, especially at low signal-to-noise ratios. Furthermore a stochastic method for the estimation of the error probability in discrimination tasks is presented. Initial experiments suggest that this method provides a valid means for the prediction of localization performance.

REFERENCES

[1] Hebrank, J.; Wright, D. Spectral cues used in the localization of sound sources on the median plane. The Journal of the Acoustical Society of America, 56 (6), 1974, 1829-1834.

[2] Wightman, F. L.; Kistler, D. J. The dominant role of low-frequency interaural time differences in sound localization. The Journal of the Acoustical Society of America, 91 (3), 1992, 1648-1661.

[3] Blauert, J. Spatial Hearing: The Psychophysics of Human Sound Localization, The MIT Press, Cambridge (USA), Revised edition, 1996 .

[4] Baumgartner, R.; Majdak, P.; Laback, B. Modeling sound-source localization in sagittal planes for human listeners. The Journal of the Acoustical Society of America, 136 (2), 2014, 791-802.

[5] MacDonald, J. A. A localization algorithm based on head-related transfer functions. The Journal of the Acoustical Society of America, 123 (6), 2008, 4290-4296.

[6] Algazi, V. R.; Duda, R. O.; Thompson D. M.; Avendano, C. The CIPIC HRTF Database. Proceedings 2001 IEEE Workshop on Applications of Signal Processing to Audio and Electroacoustics, Mohonk Mountain House, New Paltz, NY, October 21-24, 2001, 99-102.