ICA 2022
24th INTERNATIONAL CONGRESS ON ACOUSTICS
October 24(Mon) - 28(Fri), 2022 Gyeongju, Korea

PROCEEDINGS of the
24th International Congress on Acoustics

October 24 to 28, 2022 in Gyeongju, Korea

ABS-0812

Sound-source position tracking from direction-of-arrival measurements: Application to distributed first-order spherical microphone arrays

Christoph Hold[(1)], Archontis Politis[(2)], Simo Särkkä[(3)]

[(1)]Department of Signal Processing and Acoustics, Aalto University, Finland, Christoph.Hold@aalto.fi
[(2)]Department of Information Technology and Communication Sciences, Tampere University, Finland
[(3)]Department of Electrical Engineering and Automation, Aalto University, Finland

ABSTRACT
Rendering 6-degrees-of-freedom (6DoF) spatial audio requires sound-source position tracking. Without further assumptions, directional receivers, such as a spherical microphone array (SMA), can estimate the direction of arrival (DoA), but not reliably estimate sound-source distance. By utilizing multiple, distributed SMAs, further methods are available that directly infer the position in 3-D space. Typically used DoA intersection by triangulation delivers problematically noisy estimates, therefore, statistical filters are better suited. In this study, we compare the performance of different DoA to position tracking strategies. DoA angles suffer from the well-known angle wrapping problem, which is especially problematic in Gaussian filters. However, these filters are attractive due to their low computational complexity. Using circular and spherical statistics, the non-linear extensions of the Kalman filter can be formulated to explicitly treat the discontinuity of DoA angles. Furthermore, we introduce a time adaptive regularization of the filter update by the instantaneous sound-field diffuseness estimate. An experiment with three first-order SMAs in a reverberant room shows an improved distance error compared to the mean DoA intersection baseline. The results highlight the importance of treating the angle wrapping and the stabilization when incorporating the sound-field diffuseness estimate.

Keywords: Spherical Microphone Array, Position Tracking, Parametric Spatial Audio

1 INTRODUCTION

Many applications demand for tracking a target position, while the sensors only give azimuth (or bearing) and elevation measurements. Such measurements are often direction of arrival (DoA) estimates, for example emanating from wave propagation, specifically sound-waves in acoustic signal processing. An array of microphones, in particular a spherical microphone array (SMA), allows to capture directional sound and several methods have been proposed to extract the angle of incidence from such recordings. Without further assumptions, however, SMAs only allow to estimate the DoA and thus not sound-source distance. Therefore, further methods are needed to localize a source position from DoA estimates.

Utilizing multiple microphone arrays simultaneously can improve the tracking performance. Distributed SMAs allow to also estimate sound-source distance, and therefore position, for example by triangulating the respective DoA estimates. However, triangulation is known to be error prone, e. g. , because of noisy estimates, or calibration problems with multiple sensors. Particularly in acoustics, room reflections can heavily influence triangulation, since the reflections interfere with the source DoA measurement, hence leading to incorrect estimates. A strategy to mitigate the common issues arising from such geometrical approaches is to instead utilize statistical inference. Statistical filters are a powerful method in source tracking and trajectory smoothing, which can be formulated to infer a sound-source position estimate by processing the DoA measurements of distributed SMAs. The Kalman filter [4] is arguably the most prominent example here, due to its robust design and low complexity implementation. It has been employed in numerous target tracking applications with great success. For non-linear models, as in the present application, several extensions have been proposed. These methods typically rely on

linearization, or on sigma-point sampling, and the differences for the presented application will be investigated in this study.

A particular challenge of DoA angles is their circularity and inherent discontinuity. For example, an azimuth angle of 0 and 2π correspond to the same direction, and this behavior imposes further problems in formulating a tracking filter on the unit sphere. Several approaches treating angular measurements have been proposed, for example, the wrapped Kalman filter [15], modified coordinate systems [6], a formulation in spherical harmonics [11], or utilizing spherical statistics [1, 16, 5, 2]. The present article will apply and compare some (low-complexity) methodologies to first-order spherical microphone arrays and discuss their implications in the context of parametric spatial audio.

2 METHODS

2.1 Problem Formulation

The problem on hand requires tracking an object in 3-D space, by only observing angular DoA measurements. We will consider a target in 3-D with its position and velocity $\boldsymbol{x} = [x, y, z, \dot{x}, \dot{y}, \dot{z}]^\top$, observed by receivers r at position $\boldsymbol{p}_r \in \mathbb{R}^3$ in Cartesian coordinates $\boldsymbol{p} = [p_x, p_y, p_z]^\top$, each delivering a DoA measurement $\boldsymbol{y}_r$ in azimuth and elevation angles $\boldsymbol{\Omega} = [\phi, \theta]^\top \in \mathbb{S}^2$, where we may write the latter as a unit vector, formalized as a vector on the unit sphere manifold $\boldsymbol{u} \in \mathbb{S}^2 = \{\boldsymbol{u} \in \mathbb{R}^3 : ||\boldsymbol{u}|| = 1\}$.

The quantity of interest is the state vector at the current time step $\boldsymbol{x_k}$, which causes the observed measurements $\boldsymbol{y_k}$. Because the true target state is unknown and can only be observed through noisy measurements, the target state is modeled by a probability density function (PDF). In a Bayesian framework, where the target is considered to move as a Markov process, the posterior PDF contains all information given all past and current measurements. This framework allows to formulate an *optimal* estimator as a recursive filter that consists of a prediction, and an update/correction step [13], thereby determining the most likely $\boldsymbol{x_k}$ by statistical inference.

The state space model at time step k is expressed in form

$$\begin{aligned} \boldsymbol{x}_k &\sim p(\boldsymbol{x}_k|\boldsymbol{x}_{k-1}) \ , \\ \boldsymbol{y}_k &\sim p(\boldsymbol{y}_k|\boldsymbol{x}_k) \ . \end{aligned} \tag{1}$$

We model the state of a target with a prior density

$$p(\boldsymbol{x}) = \mathcal{N}(\boldsymbol{x}; \boldsymbol{\mu}, \boldsymbol{P}) \ , \tag{2}$$

where $\mathcal{N}(\boldsymbol{x}; \boldsymbol{\mu}, \boldsymbol{P})$ expresses a Gaussian PDF with mean $\boldsymbol{\mu}$ and covariance $\boldsymbol{P}$ evaluated at $\boldsymbol{x}$. The dynamics of the system are modeled by a constant velocity model. Because we use discrete time steps k in time intervals Δt, the model is discretized as

$$\boldsymbol{x}_k = \boldsymbol{A}\boldsymbol{x}_{k-1} + \boldsymbol{q}_{k-1} \ , \ \text{with } \boldsymbol{A} = \begin{bmatrix} 1 & 0 & 0 & \Delta t & 0 & 0 \\ 0 & 1 & 0 & 0 & \Delta t & 0 \\ 0 & 0 & 1 & 0 & 0 & \Delta t \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix} \ , \tag{3}$$

where $\boldsymbol{A} \in \mathbb{R}^{6\times 6}$ is the transition matrix and $\boldsymbol{q}$ process noise. The measurements at receiver r are modeled as

$$\boldsymbol{y}_k^r = h(\boldsymbol{x}_k, \boldsymbol{p}_r) + \boldsymbol{r}_k \ , \tag{4}$$

with the non-linear measurement function $h : \mathbb{R}^{n_x} \mapsto \mathbb{S}^2$, where n_x is the dimension of $\boldsymbol{x}$, and the measurement noise $\boldsymbol{r}_k$. The measurement function converts the target state $\boldsymbol{x}$ to the observed DoA angles $\boldsymbol{y}^r$ and is hence

dependent on the position of the receiver $\boldsymbol{p}_r$, which leads to the measurement model of

$$\begin{bmatrix} \phi_k^r \\ \theta_k^r \end{bmatrix} = \begin{bmatrix} \arctan(\frac{y_k - p_y^r}{x_k - p_x^r}) \\ \arctan(\frac{z_k - p_z^r}{\sqrt{(x_k - p_x^r)^2 + (y_k - p_y^r)^2}}) \end{bmatrix} \ . \tag{5}$$

The observations will only contain the DoA in terms of azimuth ϕ and elevation θ, per microphone array r, and h can be conveniently implemented as $\text{cart2sph}(\boldsymbol{x}_{x,y,z} - \boldsymbol{p}_r)$, which converts from Cartesian coordinates to azimuth and elevation angles. All measurements are then stacked into a single vector $\boldsymbol{y}_k$.

Uncertainty is modeled as the zero-mean Gaussian process noise $\boldsymbol{q}_{k-1} \sim N(\boldsymbol{0}, \boldsymbol{Q})$ and Gaussian measurement noise $\boldsymbol{r}_k \sim N(\boldsymbol{0}, \boldsymbol{R})$, which is assumed to be independent of the state and measurements. The process noise covariance after discretization is assumed to be related to q as

$$\boldsymbol{Q} = q \begin{bmatrix} \Delta t^3/3 & 0 & 0 & \Delta t^2/2 & 0 & 0 \\ 0 & \Delta t^3/3 & 0 & 0 & \Delta t^2/2 & 0 \\ 0 & 0 & \Delta t^3/3 & 0 & 0 & \Delta t^2/2 \\ \Delta t^2/2 & 0 & 0 & \Delta t & 0 & 0 \\ 0 & \Delta t^2/2 & 0 & 0 & \Delta t & 0 \\ 0 & 0 & \Delta t^2/2 & 0 & 0 & \Delta t \end{bmatrix} \ . \tag{6}$$

2.2 Intersection

When trying to find a potential source position the geometrical approach may set the baseline approach. In the geometrical approach, rays are casted from the receivers in the direction of their estimated DoAs. In practice, however, these rays might not all intersect in (3-D) space. As in [8], we may define the intersection as the point of minimal distance between rays instead as

$$\boldsymbol{p}_{\text{isc}} = (\boldsymbol{p}_1 + \tau_1 \boldsymbol{u}_1 + \boldsymbol{p}_2 + \tau_2 \boldsymbol{u}_2)/2 \ , \tag{7}$$

with

$$\tau_1 = \frac{(\boldsymbol{p}_2 - \boldsymbol{p}_1)^\top \boldsymbol{u}_1 + (\boldsymbol{p}_1 - \boldsymbol{p}_2)^\top \boldsymbol{u}_2 (\boldsymbol{u}_1^\top \boldsymbol{u}_2)}{1 - (\boldsymbol{u}_1^\top \boldsymbol{u}_2)^2} \ , \tag{8}$$

$$\tau_2 = \frac{(\boldsymbol{p}_1 - \boldsymbol{p}_2)^\top \boldsymbol{u}_2 + (\boldsymbol{p}_2 - \boldsymbol{p}_1)^\top \boldsymbol{u}_1 (\boldsymbol{u}_1^\top \boldsymbol{u}_2)}{1 - (\boldsymbol{u}_1^\top \boldsymbol{u}_2)^2} \ , \tag{9}$$

between two receivers at $\boldsymbol{p}_1$ and $\boldsymbol{p}_2$, with their respective unit vector DoA estimates $\boldsymbol{u}$. In practice, both τ are required to be positive values in order to produce an intersection in the same half-plane. The mean of all intersections from SMA receiver pairs is used as the baseline approach in this study.

2.3 Gaussian Filters

We consider a Gaussian filtering distribution

$$p(\boldsymbol{x}_k | \boldsymbol{y}_{1:k}) \simeq N(\boldsymbol{x}_k | \boldsymbol{m}_k, \boldsymbol{P}_k) \ . \tag{10}$$

The filter prediction step can be described in matrix form due to the linear transition model, leading to predicted $\boldsymbol{m}_k^-$ and $\boldsymbol{P}_k^-$ by

$$\begin{aligned} \boldsymbol{m}_k^- &= \boldsymbol{A} \boldsymbol{m}_{k-1} \ , \\ \boldsymbol{P}_k^- &= \boldsymbol{A} \boldsymbol{P}_{k-1} \boldsymbol{A}^\top + \boldsymbol{Q}_{k-1} \ . \end{aligned} \tag{11}$$

The update step involves the non-linear measurement model function h, hence requires extension strategies discussed in the following sections.

As highlighted before, angular measurements call for special care when calculating their mean and difference, hence we adapt the classical filter solutions in the following. Angular means and differences occur in multiple filtering equations, for example when calculating the difference between the predicted and the observed state, also referred to as filter innovation. We will expect problems near the wrapping boundaries, since e. g. , the angle $\phi = 0$ and $\phi = 2\pi$ represent the same angle, therefore, the mean and angular difference needs to reflect this property. One simple mitigation of the wrapping problem is to calculate the difference between two angles from arguments of their respective complex numbers [7, Ch. 2]. As any complex number can be represented as radius r and angle ϕ, with the imaginary number $i = \sqrt{-1}$, their difference becomes

$$\phi_1 - \phi_2 = \angle(\exp(i(\phi_1 - \phi_2))) \ , \tag{12}$$

which preserves direction. The (weighted) mean can be defined as the weighted mean of their complex number representations $z_i = \exp(i\phi_i)$

$$\bar{\phi} = \angle(\bar{z}) = \angle(\sum w_i z_i) \ . \tag{13}$$

The above can easily be adapted to treat the wrapping problem in $\mathbb{S}^2$, or similarly, average the components of unit vectors $\boldsymbol{u}$ and transform back to $\boldsymbol{y}$ (see also [1]).

2.3.1 Extended Kalman Filter

A natural choice for non-linear Gaussian filtering is the extended Kalman filter (EKF) that is based on a local linearization of the model by Taylor series expansion. For a linear approximation, the first two terms are sufficient, hence differentiation stops after the Jacobian. The matrix form allows for a very efficient implementation, however, the filter requires the analytical derivation of the Jacobian. For the current problem, the Jacobian (where it exists) was implemented with entries

$$\begin{aligned}
\frac{d\phi}{dx} &= \frac{-1}{1+(\frac{y-p_y}{x-p_x})^2}\frac{y-p_y}{(x-p_x)^2} \ , & \frac{d\theta}{dx} &= \frac{1}{1+\left(\frac{z-p_z}{\sqrt{(x-p_x)^2+(y-p_y)^2}}\right)^2}(-x)\frac{z-p_z}{((x-p_x)^2+(y-p_y)^2)^{3/2}} \ , \\
\frac{d\phi}{dy} &= \frac{1}{1+(\frac{y-p_y}{x-p_x})^2}\frac{1}{(x-p_x)} \ , & \frac{d\theta}{dy} &= \frac{1}{1+\left(\frac{z-p_z}{\sqrt{(x-p_x)^2+(y-p_y)^2}}\right)^2}(-y)\frac{z-p_z}{((x-p_x)^2+(y-p_y)^2)^{3/2}} \ , \\
\frac{d\phi}{dz} &= 0 \ , & \frac{d\theta}{dz} &= \frac{1}{1+\left(\frac{z-p_z}{\sqrt{(x-p_x)^2+(y-p_y)^2}}\right)^2}\frac{1}{((x-p_x)^2+(y-p_y)^2)^{1/2}} \ .
\end{aligned} \tag{14}$$

The filtering equations include the prediction step according to Eq. (11). The update step leads to the posterior with $\boldsymbol{m}_k$ and $\boldsymbol{P}_k$, using the Jacobian $\boldsymbol{H}(\cdot)$, according to [13, Alg. 5.4] as

$$\begin{aligned}
\boldsymbol{S}_k &= \boldsymbol{H}(\boldsymbol{m}_k^-)\boldsymbol{P}_k^-\boldsymbol{H}^\top(\boldsymbol{m}_k^-) + \boldsymbol{R}_k \ , \\
\boldsymbol{K}_k &= \boldsymbol{P}_k^-\boldsymbol{H}_k^\top(\boldsymbol{m}_k^-)\boldsymbol{S}_k^{-1} \ , \\
\boldsymbol{v}_k &= \boldsymbol{y}_k - h(\boldsymbol{m}_k^-) \ \text{, subject to Eq. (12)} \ , \\
\boldsymbol{m}_k &= \boldsymbol{m}_k^- + \boldsymbol{K}_k\boldsymbol{v}_k \ , \\
\boldsymbol{P}_k &= \boldsymbol{P}_k^- - \boldsymbol{K}_k\boldsymbol{S}_k\boldsymbol{K}_k^\top \ .
\end{aligned} \tag{15}$$

Note the angular difference in calculating the innovation $\boldsymbol{v}$, wherefore we call this modification EKF SPH in the following.

2.3.2 Unscented Kalman Filter

A common criticism about the extended Kalman filter is the insufficient approximation of the non-linearity and hence inferior performance in some circumstances [14]. Therefore, the slightly more flexible unscented Kalman filter (UKF) has been proposed [18], where the Gaussian is described by a set of sigma-points $\mathscr{X}$. These sigma-points can then be propagated through any (non-linear) model function g as $\mathscr{Y} = g(\mathscr{X})$. The sigma-points sample around the mean and are chosen such that they can approximate a Gaussian distribution by its mean and covariance [13, eq. 5.76] with

$$\begin{aligned} \mathbb{E}[g(\boldsymbol{x})] \simeq \boldsymbol{\mu}_U &= \sum_{i=0}^{2n} W_i^{(m)} \mathscr{Y}_i \text{ , subject to Eq. (13) ,} \\ \mathbb{C}[g(\boldsymbol{x})] \simeq \boldsymbol{S}_U &= \sum_{i=0}^{2n} W_i^{(c)} (\mathscr{Y}_i - \boldsymbol{\mu}_U)(\mathscr{Y}_i - \boldsymbol{\mu}_U)^\top \text{subject to Eq. (12) ,} \end{aligned} \tag{16}$$

reflecting the angular wrapping, similar to [1], labeled UKF SPH in the following. The sigma-points are found as [13, eq. 5.74]

$$\begin{aligned} \mathscr{X}^0 &= \boldsymbol{m} \ , \\ \mathscr{X}^i &= \boldsymbol{m} \pm \sqrt{n+\lambda}\,[\sqrt{\boldsymbol{P}}]_i \ , \end{aligned} \tag{17}$$

where the associated weights W_i are given in [13, eq. 5.77] and sum to unity, and with the square-root of the covariance matrix $\sqrt{\boldsymbol{P}}$ (e. g. , by Cholesky factorization). The parameter λ is chosen as in [13, eq. 5.75]

$$\lambda = \alpha^2 (n + \kappa) - n \ , \tag{18}$$

where $\alpha = \kappa = 1$ were set without further optimization. A deterministic sampling scheme on the unit hypersphere has been proposed in [5].

The filter prediction is again given by Eq. 11. The sigma-point sampling is only necessary for the non-linearity of the filter update, propagating the predicted state through the measurement function h. The exact filtering equations are given in [13, Alg. 5.14], however, the angular difference and weighted angular means are here subject to Eq. 12 and Eq. 13, respectively. The posterior of the update is carried out again as

$$\begin{aligned} \boldsymbol{m}_k &= \boldsymbol{m}_k^- + \boldsymbol{K}_k \boldsymbol{v}_k \ , \\ \boldsymbol{P}_k &= \boldsymbol{P}_k^- - \boldsymbol{K}_k \boldsymbol{S}_k \boldsymbol{K}_k^\top \ . \end{aligned} \tag{19}$$

A square-root form can be found in [17], which requires fewer computational operations, besides numerical stability benefits.

2.4 Spherical Distribution Filter Variant

The von Mises-Fisher (vMF) distribution properly defines a distribution on the sphere, with mean $\boldsymbol{\mu}$ and concentration parameter κ. The concentration parameter is inversely proportional to the variance. The distribution is derived by conditioning a Gaussian PDF on the hypershere, and can therefore be seen as an *intrinsic* approach to the specific characteristics of angular measurements. In contrast to a Gaussian distribution in azimuth and elevation, the vMF shows no angular stretching for increasing elevation angles (see [2]).

Let $\boldsymbol{u}$ again be the unit vector on the sphere, i. e. , $n = 3$, the probability density function of the vMF distribution is given in [7, Eq. 9.3.4]

$$f_{\text{vMF}}(\boldsymbol{u}; \boldsymbol{\mu}, \kappa) = C_3 \exp\left(\kappa \boldsymbol{\mu}^\top \boldsymbol{u}\right) \ , \tag{20}$$

where for the present case the normalising constant simplifies to $C_3 = \frac{\kappa}{\sinh \kappa}$. The distribution is uniform for $\kappa = 0$ and unimodal for $\kappa > 0$ with mean

$$\mathbb{E}[\boldsymbol{u}] = A_n(\kappa) \boldsymbol{\mu} \ , \tag{21}$$

and covariance

$$\mathbb{C}[\boldsymbol{u}] = \frac{A_n(\kappa)}{\kappa}\boldsymbol{I}_n + \left[1 - A_n^2(\kappa) - n\frac{A_n(\kappa)}{\kappa}\right]\boldsymbol{\mu}\boldsymbol{\mu}^\top \ . \tag{22}$$

For the present case of $n = 3$ the above simplifies with [7, Eq. 9.3.9]

$$A_3 = \coth\kappa - \frac{1}{\kappa} \ . \tag{23}$$

Based on the vMF distribution, a Gaussian filter has been formulated for position tracking with DoA measurements, which lie in a $\mathbb{S}^{n-1}$ manifold [2]. While the prediction step follows the standard Gaussian filter solution, they have presented a solution for the update step using sigma points as in the UKF, but modeling the measurements using vMF distribution, which is abbreviated as UKF vMF in this study. We re-arranged their solution to have access to the Kalman gain matrix $\boldsymbol{K}$, such that the update is available as in the form of Eq. (19). It should be noted that the authors mention improved performance for an iterative optimization of $\boldsymbol{A}$ and $\boldsymbol{R}$, which was not carried out in favor for simplicity and comparability to the other presented methods.

2.5 Parameter Extraction

Considering a basic sound-field model of a sound-source in free-field conditions, the emitted sound impinges at the microphone as plane-waves. The DoA of a sound-source is in opposite direction of its net acoustic energy flow (i.e. acoustic intensity, or dependent on definition in the same direction). Extracting this intensity vector $\boldsymbol{i}$ is particularly convenient for spherical microphone arrays, as the spherical harmonic (SH) expansion of a sound-field up to first order is proportional to the pressure p (zeroth order), and to the pressure gradient $\boldsymbol{v}$ (first order). The (pseudo-) intensity vector is proportional to the measured

$$\boldsymbol{i} \propto \Re\{p^\mathsf{H}\boldsymbol{v}\} \ . \tag{24}$$

The vector direction directly estimates the DoA Ω^{DoA} of its predominant signal component

$$\Omega^{\mathrm{DoA}} = \angle\boldsymbol{i} \ . \tag{25}$$

The measured azimuth and elevation were extracted per time sample and then averaged per processing block over Δt, and all simultaneous measurements stacked into $\boldsymbol{y}_k$. It has been shown that this simple technique leads to reliable DoA estimates from first-order SH components [3]. However, reflections can influence the extracted direction, because typically reflections are linearly correlated to the sound-source signal, but from competing DoAs.

The estimates p and $\boldsymbol{v}$ also deliver an estimate of the sound-field sector energy E and diffuseness parameter ψ

$$\psi = 1 - \frac{||\boldsymbol{i}||}{E} = 1 - \frac{||\boldsymbol{i}||}{|p|^2 + \boldsymbol{v}^\mathsf{H}\boldsymbol{v}} \tag{26}$$

as an indicator of the degree of deviation from a purely propagating soundfield to fully reactive or isotropic sound fields [12]. The diffuseness ψ is related to the length of the active intensity vector and is defined in $\psi \in [0,1]$, which results in $\psi = 0$ for a single impinging plane-wave or far-field source, and $\psi = 1$ for no observed net flow, occurring for example in a fully diffuse sound-field, such as dense reverberation. It can be interpreted as the directionality of the sound-field intensity flow, which we will explore as a measure of the *reliability* of the DoA estimate. The intuition here is that in a dry environment, the DoA is dominated by the direct path net intensity flow of the sound-source of interest, whereas in a reverberant environment the superposition with multiple reflections will result in a less reliable estimate. A similar and more elaborate concept is the direct path dominance (DPD) test [10], applied to the pseudo-intensity vector in [9], which separates the covariance matrix into sub-spaces instead and is therefore computationally more demanding.

2.6 Modification of Filter Update

Under the assumption that a highly diffuse sound-field results in less reliable sound-source direction estimates, the tracking algorithm may be further optimized for spatial audio applications. The sound-field diffuseness estimate

measured by each spherical microphone array can function as an indicator for unfavorable DoA estimation conditions. The estimated sound-field diffuseness value seems to be a promising parameter in order to incorporate additional sound-field information into the filtering algorithm. The diffuseness value does furthermore not only indicate an unreliable measurement due to reverberation, it also goes to one if a DoA estimation is not possible due to a lack of input signal. It therefore constitutes a threshold independent measure to detect insufficient input signal, as it may occur in a speaker pause. A moving sound-source is likely to continue moving during a short break (i. e., in between words of a moving speaker), wherefore loosely continuing the dynamics seem sensible.

The filter update is regularized under unreliable conditions, which are indicated by high diffuseness values. This work introduces a simple thresholding approach on the measurement noise $\boldsymbol{R}$, influencing entry R_r of any effected receiver r as

$$R_r^{\text{mod}} = \tau_r R_r \ \text{, if } \ \psi_r > \psi_{\text{TH}} \ , \tag{27}$$

and updates $\boldsymbol{S}$ accordingly in the form of

$$\boldsymbol{S}_k = \boldsymbol{S}_k + \boldsymbol{R}_k^{\text{mod}} \ . \tag{28}$$

Because the Kalman gain $\boldsymbol{K}$ is inversely proportional to $\boldsymbol{S}$, the update

$$\begin{aligned} \boldsymbol{m}_k &= \boldsymbol{m}_k^- + \boldsymbol{K}_k \boldsymbol{v}_k \ , \\ \boldsymbol{P}_k &= \boldsymbol{P}_k^- - \boldsymbol{K}_k \boldsymbol{S}_k \boldsymbol{K}_k^T \end{aligned} \tag{29}$$

heavily favors the predicted solution $\boldsymbol{m}_k^-$ over the innovation $\boldsymbol{v}_k$ in case of high diffuseness estimates. This strategy avoids manipulating the Kalman gain directly, which might result in a inconsistent filtering formulation. We tested a threshold of $\psi_{\text{TH}} = 0.1$ and modulated with the corresponding diffuseness value as $\tau_r = 10\psi_r$. Note that these values are chosen heuristically and further strategies should be investigated.

3 EVALUATION AND DISCUSSION

The evaluation of the presented methodology was carried out on simulated recordings of a moving source in a reverberant room.[1] An image-source reverberation model with a reverb time of $RT_{60} = 0.5\,\text{s}$ calculated the reflection pattern of the room outlined in Fig. 1 and Fig. 2. The impulse response at the microphones was updated with the moving source in 0.2 m increments. The sound-source was moving along a trajectory of 16 m with a velocity of 1 m/s in a setup shown in Fig. 1. The sound-source first moves along a typical movement for speech along the horizontal plane, and then corners sharply into an ascending motion. The virtual sound-source emitted a white noise sequence, band-passed between 100 Hz to 10 kHz. The virtual microphone arrays delivered a set of first-order spherical harmonic audio signals, which were then split into blocks of 1024 samples, followed by the parameter extraction detailed in Sec. 2.5. Virtual first-order spherical harmonic receivers captured the scenario in two different typical arrangements. First, an arrangement of three microphone arrays that captured the room from various angles was simulated, depicted in Fig. 1. It is noted here, that the method does not require the receivers to enclose the trajectory of the target, which is demonstrated by the target trajectory leaving the visualized triangulation. Second, an arrangement consisting of three microphone arrays spaced along a common axis, similar to a linear array, which is particularly relevant in practical applications. This arrangement shown in Fig. 2 demonstrates the performance with all receivers located at the same height, besides exemplifying the azimuthal angle wrapping challenge.

These scenarios were chosen to demonstrate multiple effects of the filter designs. We expect critical performance differences around the circularity/ angle discontinuity, since the source passes the $\pm\pi$ relative azimuth angle for the first part of the trajectory, i. e., for microphone m1 in scene 1 and for all three microphones in scene 2. The sharp corner should then uncover problems in the adaption of the filters. Furthermore, the increasing elevation in the second part of the trajectory challenges the assumptions, as only the vMF filter models the measurement statistics correctly in this case. The tracker was initialized with $\boldsymbol{m}_0$ offset from the true value by a realization of standard normal noise, in order to investigate the convergence of the algorithms. The measurement noise was set to a diagonal matrix where the entries correspond to an uncertainty of 5°, and the process noise to $\boldsymbol{P}_0 = \boldsymbol{I}$.

[1] implementation using : https://github.com/polarch/shoebox-roomsim

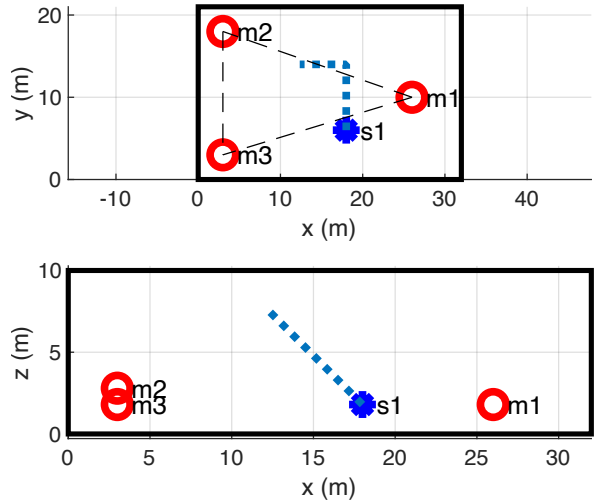

Figure 1. Simulated Scene 1 in two cross-sections, m marks the virtual spherical microphone arrays, s the simulated source moving on the dashed line. The simulated room geometry is indicated by the solid boundary.

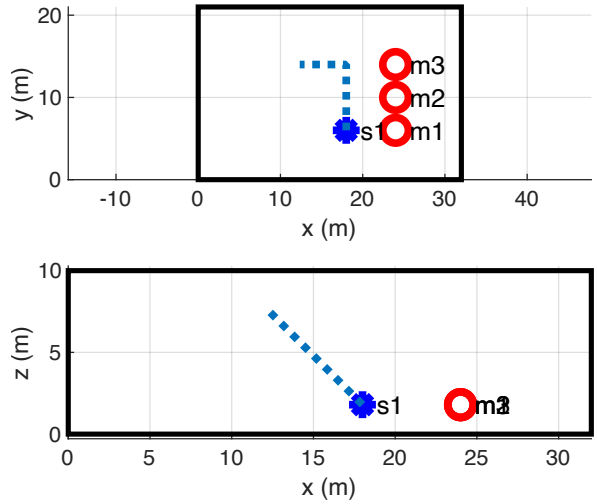

Figure 2. Simulated Scene 2 in two cross-sections, m marks the virtual spherical microphone arrays, s the simulated source moving on the dashed line. The simulated room geometry is indicated by the solid boundary.

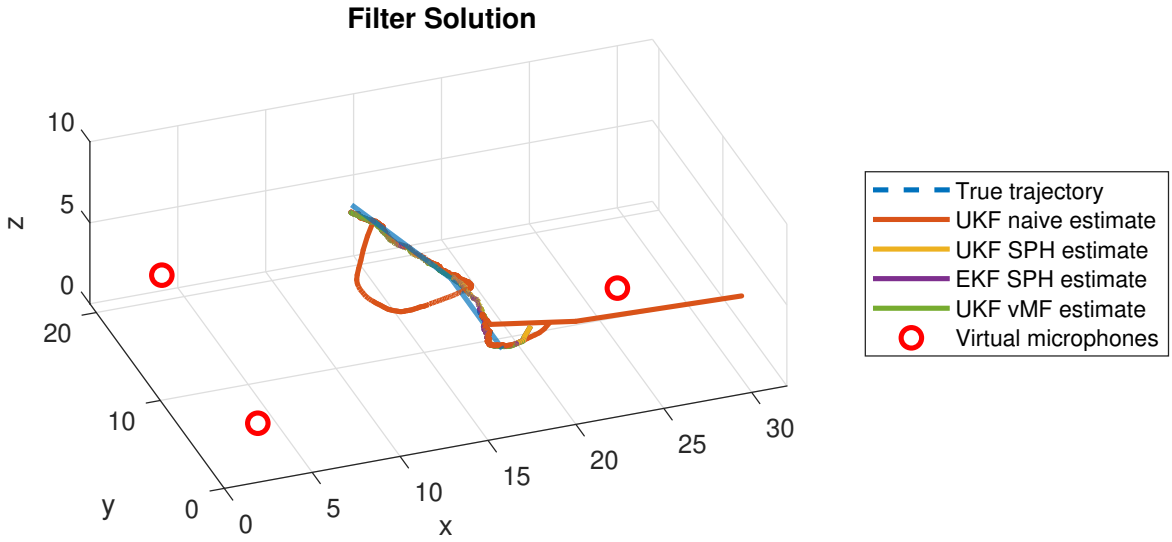

Figure 3. Scene 1 filter estimates in comparison to the true trajectory. The room geometry is depicted as the coordinate limits. The virtual microphone array locations are marked as circles.

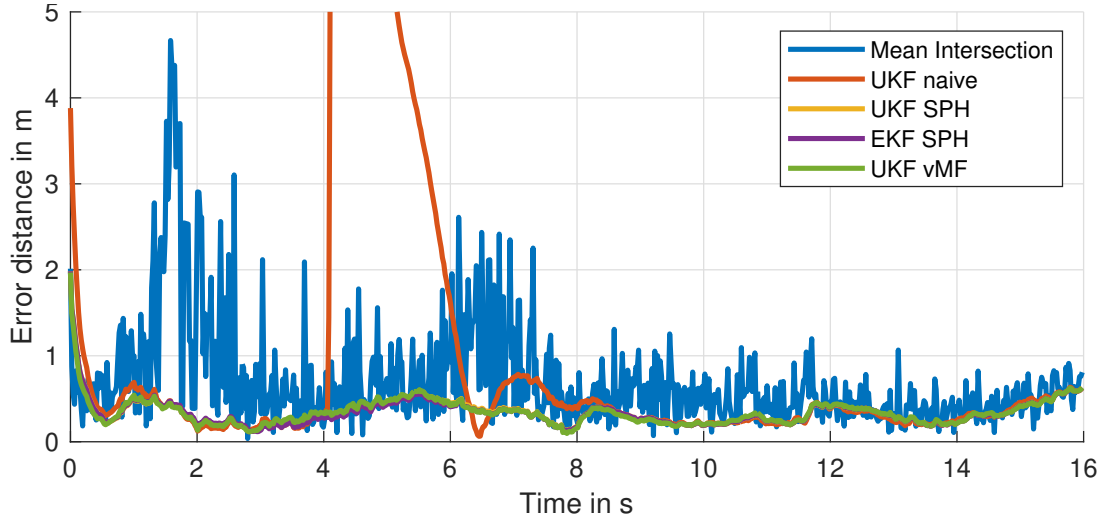

Figure 4. Position error distance of the evaluated estimators in comparison to the true trajectory for scene 1.

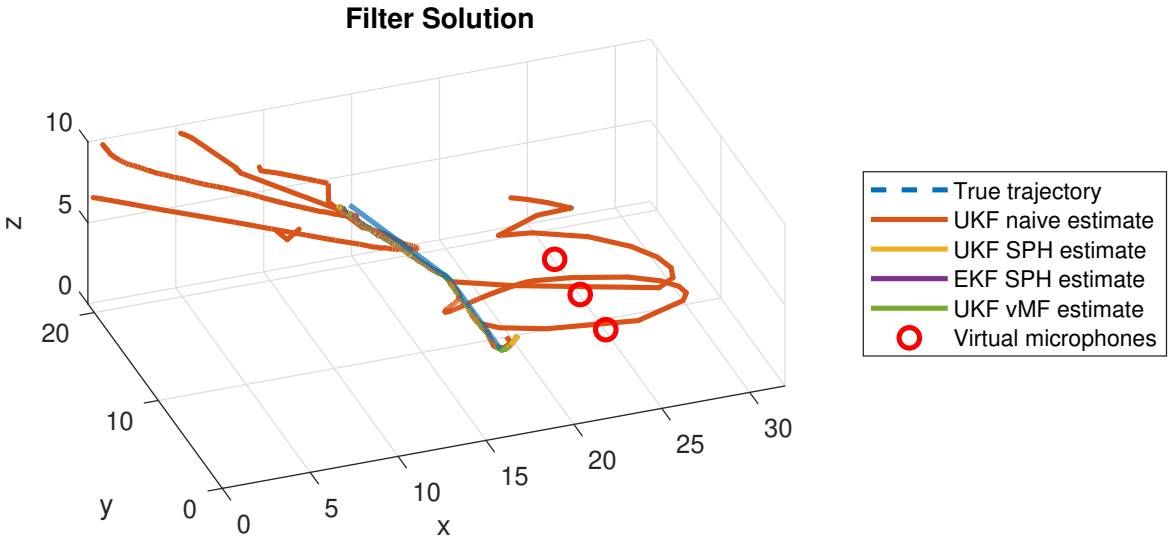

Figure 5. Filter estimates in comparison to the true trajectory. The room geometry is depicted as the coordinate limits. The virtual microphone array locations are marked as circles.

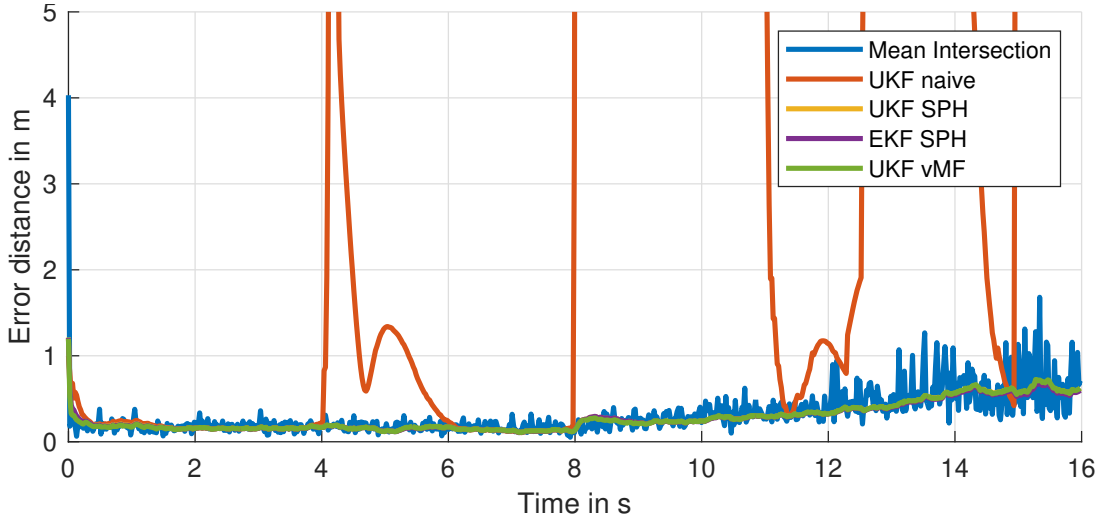

Figure 6. Position error distance of the evaluated estimators in comparison to the true trajectory for scene 2.

Table 1. Position error distance (RMSE) of filter estimate to true trajectory.

Estimator	RMSE Scene 1	RMSE Scene 2
Mean Intersections	0.9322	0.4066
UKF naive	5.9020	12.4264
UKF SPH	0.3749	0.3247
EKF SPH	0.3694	0.3179
UKF vMF	0.3716	0.3240

Summarizing the performance, Tab. 1 shows the evaluated root-mean-square error (RMSE) Euclidean distance between the true trajectory and the filter estimates. The filter performance results are similar and consistent between both scenes. All solutions using angular filtering produce nearly identical results for the presented scenes. In comparison to the naive UKF implementation (i. e. , without adapting the filtering equations), the spherical variants show a clear improvement. Filters UKF SPH and EKF SPH use the spherical statistics Eq. 12 and 13, whereas UKF vMF utilizes the vMF distribution. Furthermore, the spherical filters improve on the geometric intersection approach baseline, with also much smoother and more stable results.

Figure 3 shows the tracking filter estimation in comparison with the simulated true trajectory for scene 1. It shows that all algorithms, except the naive UKF, perform equally well and deliver accurate tracking of the sound-source. Figure 4 visualizes the distance RMSE over time, indicating that all filters converge quickly from the intentional initialization offset. The results show a clear indication that neglecting the circular nature of spherical angle measurements leads to poor performance, which is consistent with the literature, e. g. , [1, 2] The naive UKF without angular measurement adjustments to the filtering algorithm, leads to a significant estimation error and trajectory divergence just at the point where SMA $m1$ produces measurements around $\pm\pi$ azimuth, which also the information of two additional SMAs can not counterbalance. After the detour, the filter converges again towards the solution of the other estimators. Scene 2 seems to highlight these problems even more significantly, as demonstrated in Fig. 5 and 6. Again, the naive UKF solution produces unacceptable estimation errors, whenever the trajectory passes the $\pm\pi$ azimuth wrapping of each receiver. The estimation error generally increased for increasing elevation. With all receivers on one axis, the relative differences decrease, hence, statistical inference becomes harder. Additionally, the measured diffuseness increased here, which led to higher uncertainty in the DoA measurements according to Fig. 7. Interestingly, UKF vMF could only show a marginal improvement over UKF SPH, which came at a significant increase in algorithmic complexity.

The estimated sound-field diffuseness value seems to be a promising parameter in order to incorporate additional sound-field information into the filtering algorithm. This is particularly interesting, as these parameters are usually extracted in parametric spatial audio at a very low computational cost. The diffuseness value does furthermore not only indicate an unreliable measurement due to reverberation, it also reacts when DoA estimation is not possible due to a lack of input signal. It therefore constitutes a threshold independent measure to detect insufficient input, as apparent from Fig.7 for the first few time blocks. Future work could extend to multi-source algorithms and exploring dedicated subspace-methods, e. g. , for the parameter estimation.

4 CONCLUSIONS

For this study, multiple computationally efficient non-linear Kalman filters were explored for 3-D target tracking from DoA measurements. This concept was then applied to a moving source in a reverberant room, resembling a both practical and challenging scenario. Concepts of addressing the particularities of spherical measurements were demonstrated and evaluated. In order to mitigate problematic behavior in low SNR scenarios, additional filter regularization dependent on the estimated sound-field diffuseness was explored. The concept of augmenting the classical Kalman filter with information available in parametric spatial audio seems a simple and promising

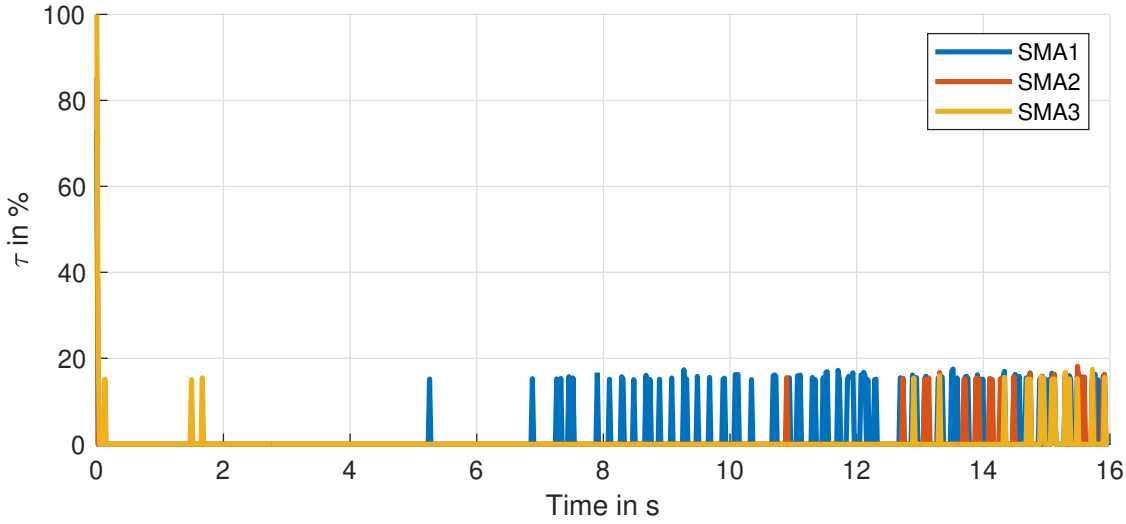

Figure 7. Filter update modifier τ in percent, shown for UKF SPH in scene 2.

strategy in order to optimize sound-source position tracking. [2]

REFERENCES

[1] D. F. Crouse. Cubature/unscented/sigma point Kalman filtering with angular measurement models. *2015 18th International Conference on Information Fusion, Fusion 2015*, pages 1550–1557, 2015.

[2] A. F. Garcia-Fernandez, F. Tronarp, and S. Särkkä. Gaussian target tracking with direction-of-arrival von mises-fisher measurements. *IEEE Transactions on Signal Processing*, 67(11):2960–2972, 2019.

[3] D. P. Jarrett, E. A. Habets, and P. A. Naylor. 3D Source localization in the spherical harmonic domain using a pseudointensity vector. In *European Signal Processing Conference*, 2010.

[4] R. E. Kalman. A new approach to linear filtering and prediction problems. *Transactions of the ASME–Journal of Basic Engineering*, 82(Series D):35–45, 1960.

[5] G. Kurz, I. Gilitschenski, and U. D. Hanebeck. Unscented von mises-fisher filtering. *IEEE Signal Processing Letters*, 23(4):463–467, 2016.

[6] M. Mallick, L. Mihaylova, S. Arulampalam, and Y. Yan. Angle-only filtering in 3D using modified spherical and log spherical coordinates. *Fusion 2011 - 14th International Conference on Information Fusion*, 2011.

[7] K. V. Mardia and P. E. Jupp. *Directional Statistics*. Wiley, 2000.

[8] L. McCormack, A. Politis, T. McKenzie, C. Hold, and V. Pulkki. Object-Based Six-Degrees-of-Freedom Rendering of Sound Scenes Captured with Multiple Ambisonic Receivers. *AES: Journal of the Audio Engineering Society*, 70(5):355–372, may 2022.

[9] A. H. Moore, C. Evers, P. A. Naylor, D. L. Alon, and B. Rafaely. Direction of arrival estimation using pseudo-intensity vectors with direct-path dominance test. *2015 23rd European Signal Processing Conference, EUSIPCO 2015*, pages 2296–2300, 2015.

[2] Implementation available online https://github.com/chris-hld/doa2pos.

[10] O. Nadiri and B. Rafaely. Localization of multiple speakers under high reverberation using a spherical microphone array and the direct-path dominance test. *IEEE Transactions on Audio, Speech and Language Processing*, 22(10):1494–1505, 2014.

[11] F. Pfaff, G. Kurz, and U. D. Hanebeck. Filtering on the unit sphere using spherical harmonics. *IEEE International Conference on Multisensor Fusion and Integration for Intelligent Systems*, 2017-Novem:124–130, 2017.

[12] V. Pulkki, S. Delikaris-Manias, and A. Politis. *Parametric Time – Frequency Domain Spatial Audio*. Wiley, 2018.

[13] S. Särkkä. *Bayesian filtering and smoothing*. Cambridge University Press, Cambridge, 2013.

[14] J. Shen, Y. Liu, S. Wang, and Z. Sun. Evaluation of unscented Kalman filter and extended Kalman filter for radar tracking data filtering. *Proceedings - UKSim-AMSS 8th European Modelling Symposium on Computer Modelling and Simulation, EMS 2014*, pages 190–194, 2014.

[15] J. Traa and P. Smaragdis. A wrapped kalman filter for azimuthal speaker tracking. *IEEE Signal Processing Letters*, 20(12):1257–1260, 2013.

[16] J. Traa and P. Smaragdis. Multiple speaker tracking with the Factorial von Mises-Fisher Filter. In *2014 IEEE International Workshop on Machine Learning for Signal Processing (MLSP)*, number 3, pages 1–6. IEEE, sep 2014.

[17] R. Van Der Merwe and E. A. Wan. The square-root unscented Kalman filter for state and parameter-estimation. *ICASSP, IEEE International Conference on Acoustics, Speech and Signal Processing - Proceedings*, 6:3461–3464, 2001.

[18] E. A. Wan and R. Van Der Merwe. The unscented Kalman filter for nonlinear estimation. *IEEE 2000 Adaptive Systems for Signal Processing, Communications, and Control Symposium, AS-SPCC 2000*, pages 153–158, 2000.