A Wrapped Kalman Filter for Azimuthal Speaker Tracking

Johannes Traa, *Student Member, IEEE,* and Paris Smaragdis, *Member, IEEE*

***Abstract*—We present the wrapped Kalman filter (WKF) for tracking the azimuth of a speaker with a compact, 3-channel microphone array. Traditional extended and unscented filters assume that the observation is a rotating vector in $\mathbb{R}^2$. However, the azimuth inhabits a 1D subspace: the unit circle. We model the state variable with a wrapped Gaussian distribution and show that this achieves a lower mean squared error than 2D methods. We demonstrate the superior tracking performance of the WKF in simulated and real reverberant environments.**

***Index Terms*—Kalman filter, wrapped Gaussian, circular statistics, direction of arrival tracking**

I. Introduction

Array-based acoustic source localization and tracking is an important area of research, particularly for speech-driven interfaces [1], [2], [3]. These technologies often require a smooth estimate of the source position over time. Traditionally, the Kalman filter (KF) [4] and its extended (EKF) [5] and unscented (UKF) [6] variants provide this capability. The KF is only applicable to linear-Gaussian dynamical systems. However, it can be generalized to non-linear models by first-order linearization (in the EKF) or direct representation of the non-linearities through the unscented transform (in the UKF).

In this paper, we are interested in tracking an acoustic source with a compact array. Although range information cannot be accurately estimated in this case, we can track the source's direction-of-arrival (DOA). Thus, the state position is a circular variable, i.e. it is constrained to lie on the unit circle. One popular technique is to represent it as a scalar in $\mathbb{R}^1$ and the observation as a rotating vector in $\mathbb{R}^2$ [7]. Unfortunately, this 2D model introduces additional noise to the 1D system.

We present the *wrapped Kalman filter* (WKF) as an alternative to the EKF and UKF for recursive Bayesian inference of a circular variable. The WKF tracks the hidden state of a wrapped dynamical system whose 1D observations lie in $[-\pi, \pi]$. It maintains an estimate of the posterior state distribution $P(\theta_t|y_{1:t})$ by approximating it as a wrapped Gaussian (WG) [8]. The resulting filter equations involve a weighted sum of infinitely many innovation terms. Fortunately, this sum can be truncated to only a few terms in practice without a loss in performance. The same technique was used in [9] to train a wrapped hidden Markov model.

In this paper, we

- introduce a wrapped dynamical system (WDS) model
- propose the wrapped Kalman filter (WKF) to perform recursive inference for the WDS
- interpret the WKF alternatively as a KF with measurement fusion or an approximation of a switching filter
- present experimental results showing the superior performance of the WKF over extended and unscented filters for azimuthal speaker tracking

II. Wrapped Gaussian distribution

A circular random variable $\theta \in \mathbb{S}^1$ is one that lies in the range $[-\pi, \pi]$ and whose statistics are identical at the boundaries, such as phase or azimuth angle. The wrapped Gaussian (WG) distribution [8, Chapter 3] is the result of transforming a random variable $\gamma \sim \mathcal{N}\left(\mu, \sigma^2\right)$ via the mapping $\psi : \mathbb{R}^1 \to \mathbb{S}^1$:

$$\theta = \psi(\gamma) = \mathrm{mod}\left(\gamma + \pi, 2\pi\right) - \pi \quad . \tag{1}$$

Thus, the pdf is given as:

$$P(\theta\,;\,\mu,\sigma^2) = \sum_{l=-\infty}^{\infty} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(\theta-(\mu+2\pi l))^2}{2\sigma^2}} \,,\ \ \theta \in \mathbb{S}^1 \quad . \tag{2}$$

We can visualize (2) on the unit circle or in $\mathbb{S}^1$ (see Fig. 1).

III. State space models for wrapped filtering

In this section, we present two dynamical systems that model the evolution of a circular state variable.

A. Rotating vector model

The following rotating-vector state space model (RVM) [7] is often used for filtering circular data:

$$\mathbf{x}_t = \begin{bmatrix} 1 & dt \\ 0 & 1 \end{bmatrix} \mathbf{x}_{t-1} + \mathbf{v}_t \ \ , \ \ \mathbf{v}_t \sim \mathcal{N}\left(\mathbf{0}, \begin{bmatrix} \lambda_{v,1}^2 & 0 \\ 0 & \lambda_{v,2}^2 \end{bmatrix}\right) \tag{3}$$

$$\mathbf{y}_t = \begin{bmatrix} \cos(x_{t,1}) \\ \sin(x_{t,1}) \end{bmatrix} + \mathbf{w}_t \ \ , \ \ \mathbf{w}_t \sim \mathcal{N}\left(\mathbf{0}, \lambda_w^2 \mathbf{I}\right) \quad . \tag{4}$$

The state vector $\mathbf{x}_t \in \mathbb{R}^2$ consists of position and velocity, $\mathbf{y}_t \in \mathbb{R}^2$ is the observation vector, dt is the time increment, and $\mathbf{v}_t \in \mathbb{R}^2$ and $\mathbf{w}_t \in \mathbb{R}^2$ are the process and measurement noise, respectively. We assume that the noise variances $\lambda_{v,:}^2$ and λ_w^2 are either known or can be estimated empirically. Since the measurement equation (4) involves a non-linear transformation of the state, one must resort to the EKF or UKF to infer the state sequence.

The drawback of this model is that it regards the observation as a 2D vector when the state is truly 1D (and can be inferred

Copyright (c) 2013 IEEE. Personal use of this material is permitted. However, permission to use this material for any other purposes must be obtained from the IEEE by sending a request to pubs-permissions@ieee.org.

J. Traa is a PhD student in the ECE department at the University of Illinois at Urbana-Champaign (UIUC) (e-mail: traa2@illinois.edu).

P. Smaragdis holds a joint faculty position in ECE and CS at UIUC and works with Adobe Systems, Inc. (e-mail: paris@illinois.edu)

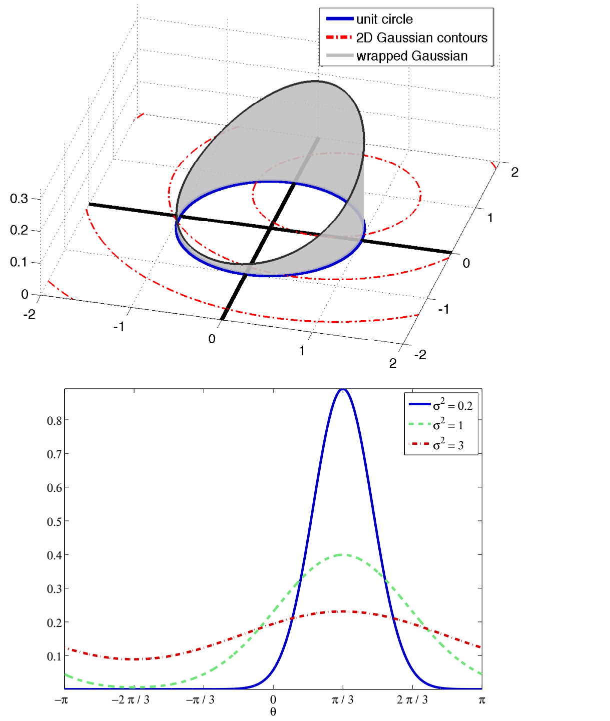

Fig. 1. (Top) Wrapped Gausian pdf ($\mu = \frac{\pi}{3}$) on the unit circle in $\mathbb{R}^2$ shown with 2D Gaussian contours ($\sigma^2 = 0.8$). (Bottom) WG pdf in $[-\pi, \pi]$ ($\mu = \frac{\pi}{3}$ and varying σ^2). The θ axis is the unit circle, unfolded.

via $\angle \mathbf{y}_t$). This introduces additional noise to the system that limits the tracking capabilities of the filters. We will show that tracking with a 1D observation model (presented next) improves tracking performance. Contours of the 2D measurement distribution are shown in the top panel of Fig. 1.

B. Wrapped dynamical system (WDS)

The position-only WDS is described by:

$$\theta_t = \psi\left(\theta_{t-1} + v_t\right) \quad , \quad v_t \sim \mathcal{N}\left(0, \sigma^2_{v,1}\right) \tag{5}$$

$$y_t = \psi\left(\theta_t + w_t\right) \quad , \quad w_t \sim \mathcal{N}(0, \sigma^2_w) \tag{6}$$

where $\theta_t, y_t \in \mathbb{S}^1$ and $v_t, w_t \in \mathbb{R}^1$. Velocity is trivially included by extending the state vector and including a noise term $\sigma^2_{v,2}$ in the state model since it is not wrapped and does not appear in (6). Thus, we will only consider position in the derivation of the WKF (for the sake of clarity) and incorporate velocity in the final algorithm. A typical sample path of the WDS is shown in Fig. 2 with an observation sequence and the WKF state estimate.

The advantage of the WDS over the RVM is that the observations are treated as 1D quantities. We can expect that filtering in this model will be more accurate since it is easier to infer the hidden state sequence $\theta_{1:T}$ from lower-dimensional measurements. Conceptually, inference in the WDS is achieved by Rao-Blackwellisation [10] of inference in the RVM with the radius of $\mathbf{y}_t$ in (4) marginalized out.

IV. Filtering for the Wrapped Dynamical System

Now we derive the wrapped Kalman filter (WKF). The filtered state distribution does not remain WG over time, but

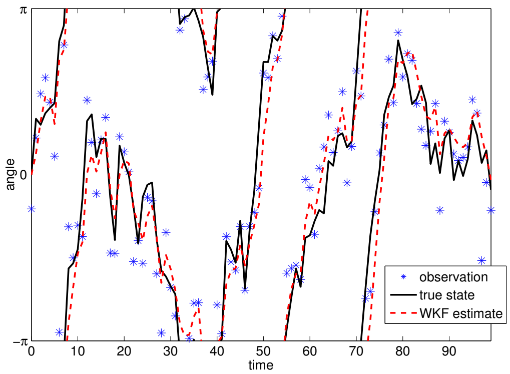

Fig. 2. Sample path and observation sequence for the wrapped dynamical system with position and velocity state components. The WKF tracks the WDS despite wrapping effects at $-\pi$ and π. ($\sigma^2_{v,1} = \sigma^2_w = 0.5, \sigma^2_{v,2} = 0.001$)

we can approximate it as such. This leads to a *correct* step with two equivalent interpretations. In one, a single WG component is updated via 2π-periodic copies of the observation (akin to measurement fusion). And in the other, all WG components are updated using a single observation. The WKF is then shown to be a good approximation of a switching Kalman filter.

A. Wrapped Kalman filtering

The WG allows us to model the wrapping function $\psi(\theta)$ in (5)-(6). The filtered state distribution at time $t-1$ is

$$P(\theta_{t-1}|y_{1:t-1}) = \sum_{l=-\infty}^{\infty} P_l(\theta_{t-1}|y_{1:t-1}) \tag{7}$$

$$= \sum_{l=-\infty}^{\infty} \mathcal{N}(\theta_{t-1}\,;\, \mu_{t-1} + 2\pi l, \sigma^2_{\theta_{t-1}}) \ . \tag{8}$$

At each step, we *predict* the next state distribution:

$$P(\theta_t|y_{1:t-1}) = \int P(\theta_t|\theta_{t-1}) P(\theta_{t-1}|y_{1:t-1})\, d\theta_{t-1} \tag{9}$$

$$= \int P(\theta_t|\theta_{t-1}) \sum_{l=-\infty}^{\infty} P_l(\theta_{t-1}|y_{1:t-1})\, d\theta_{t-1} \tag{10}$$

$$= \sum_{l=-\infty}^{\infty} \int P(\theta_t|\theta_{t-1}) P_l(\theta_{t-1}|y_{1:t-1})\, d\theta_{t-1} \tag{11}$$

$$= \sum_{l=-\infty}^{\infty} P_l(\theta_t|y_{1:t-1}) \ , \tag{12}$$

and then *correct* this prediction:

$$P(\theta_t|y_{1:t}) \propto P(y_t|\theta_t) P(\theta_t|y_{1:t-1}) \tag{13}$$

$$\propto \left[\sum_{m=-\infty}^{\infty} P_m(y_t|\theta_t)\right] \left[\sum_{l=-\infty}^{\infty} P_l(\theta_t|y_{1:t-1})\right] . \tag{14}$$

This is an exponentially-growing sum of increasingly differing Gaussian components. We approximate it at time t with a WG by considering one term of the predicted density and interpreting the observation as being replicated [11, Chapter 4]. This gives a weighted sum of 2π-periodic Gaussians:

$$\tilde{P}(\theta_t|y_{1:t}) \propto \sum_{l=-\infty}^{\infty} P(y_t + 2\pi l|\theta_t)\, P_0(\theta_t|y_{1:t-1})\, \eta_{t,l} \ , \tag{15}$$

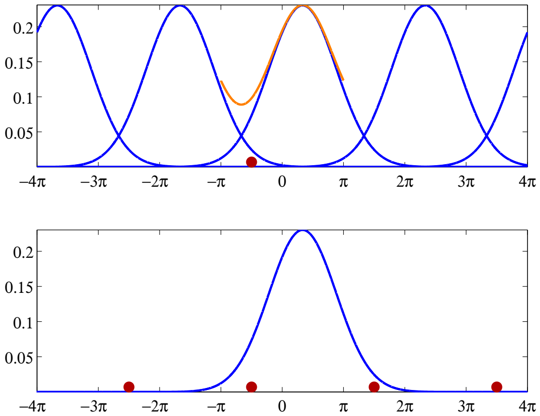

Fig. 3. Two interpretations of the *correct* step in the WKF. (Top) Single observation and periodic Gaussians (wrapped Gaussian in $[-\pi, \pi]$ overlaid). (Bottom) Single Gaussian and periodic observations. ($\mu = \frac{\pi}{3}$, $\sigma^2 = 3$)

Algorithm 1 Wrapped Kalman Filter

```
Predict
    $\widehat{\boldsymbol{\mu}}_t^- = \mathbf{A}\,\widehat{\boldsymbol{\mu}}_{t-1}$
    $\widehat{\mu}_{t,1}^- = \psi\left(\widehat{\mu}_{t,1}^-\right)$
    $\widehat{\boldsymbol{\Sigma}}_t^- = \mathbf{A}\,\widehat{\boldsymbol{\Sigma}}_{t-1}\,\mathbf{A}^T + \boldsymbol{\Sigma}_v$
Correct
    $\mathbf{K}_t = \frac{\widehat{\boldsymbol{\Sigma}}_t^-\,\mathbf{B}^T}{\mathbf{B}\,\widehat{\boldsymbol{\Sigma}}_t^-\,\mathbf{B}^T + \sigma_w^2}$
    $g_t = \sum_{l=-1}^{1} \left((y_t + 2\pi l) - \widehat{\mu}_{t,1}^-\right) \eta_{t,l}$
    $\widehat{\boldsymbol{\mu}}_t = \widehat{\boldsymbol{\mu}}_t^- + \mathbf{K}_t\, g_t$
    $\widehat{\mu}_{t,1} = \psi\left(\widehat{\mu}_{t,1}\right)$
    $\widehat{\boldsymbol{\Sigma}}_t = \left(\mathbf{I} - \mathbf{K}_t\,\mathbf{B}\right)\widehat{\boldsymbol{\Sigma}}_t^-$
```

where

$$\eta_{t,l} = \frac{\mathcal{N}(y_t + 2\pi l\,;\, \widehat{\mu}_{t,1}^-, \sigma_w^2)}{\sum_{m=-\infty}^{\infty} \mathcal{N}(y_t + 2\pi m\,;\, \widehat{\mu}_{t,1}^-, \sigma_w^2)} \quad , \tag{16}$$

represents the probability of a replicate. The posterior at time t is approximated by finding the closest Gaussian distribution to (15) via moment-matching and then repeating it every 2π. We can view this as a measurement fusion step [12]. Equivalently, we could ignore all but a single measurement term in (14) and proceed as before (see Fig. 3).

The filtering procedure is summarized in Algorithm 1. A velocity component has been incorporated. $\widehat{\boldsymbol{\mu}}_t$ and $\widehat{\boldsymbol{\Sigma}}_t$ are the estimated state mean and covariance, $\mathbf{K}_t$ is the Kalman gain, the state transition matrix $\mathbf{A}$ is the same as in (3), and the measurement matrix is a row vector $\mathbf{B} = [1, 0]$. A minus sign superscript indicates a prediction. A composite innovation g_t is formed via a weighted average of innovation terms. The position estimate must be wrapped to $\mathbb{S}^1$. We can see from Algorithm 1 that the complexity of the WKF is marginally greater than that of a conventional Kalman filter.

We truncate the WG to 3 terms in practice since we only care about wrapping effects in $[\widehat{\mu}_{t,1} - 2\pi, \widehat{\mu}_{t,1} + 2\pi]$. Thus, we only need to consider 3 replicates of y_t. For very high noise levels (e.g. $\sigma_w^2 > 2$), this degree of truncation may be inadequate. However, we cannot expect to track the state with any confidence on $\mathbb{S}^1$ in such harsh conditions, so we will ultimately only be interested in cases where 3 terms suffice.

B. WKF as an approximation of a switching Kalman filter

Modeling the state distribution as a set of locked Gaussians implies a generative model where we sample the observation from a single WG component. We can incorporate a hidden indicator variable z_t that selects what component is active at time t. The result is that we have a switching measurement equation. The state is a vector $\boldsymbol{\theta}_t$ of the WG component means and the observation y_t is a selected mean plus noise:

$$\boldsymbol{\theta}_t = \boldsymbol{\theta}_{t-1} + \mathbf{v}_t \quad , \quad \mathbf{v}_t \sim \mathcal{N}(\mathbf{0}, \sigma_v^2 \mathbf{J}) \quad , \tag{17}$$

$$y_t = \mathbf{B}_{z_t}\,\boldsymbol{\theta}_t + w_t \quad , \quad w_t \sim \mathcal{N}(0, \sigma_w^2) \quad , \tag{18}$$

where $\mathbf{J}$ is a ones matrix and z_t selects a measurement matrix:

$$\mathbf{B}_{z_t} = [\ldots, 0, 0, 1, 0, 0, \ldots] \quad . \tag{19}$$

The state distribution has a rank-1 covariance matrix $\boldsymbol{\Sigma}_t = \sigma_t^2 \mathbf{J}$ because the WG means (components of $\boldsymbol{\theta}_t$) are locked. This leads to a standard switching Kalman filter (SKF) [13]. The WKF is equivalent to the SKF when the transition distribution $P(z_t|z_{t-1})$ is modeled as uniform. We found that this is not a significant drawback as the weights $\eta_{t,l}$ are sufficient to capture transitions between neighboring WG components. Forming the composite innovation g_t is analogous to the "collapse" operation of the SKF.

V. Experiments

We now show that the WKF can track an acoustic source more accurately than either the EKF or UKF.

A. Array and DOA estimator

The array used in the following experiments consists of three omnidirectional microphones placed in a right triangle configuration (the perpendicular sides are 1 cm in length).

DOA estimates are extracted from the audio stream to use as measurements. We sample the audio at 16,000 Hz and segment it into windows of 1,024 samples with 3/4 overlap. The DFT of each window is calculated for each channel and inter-channel time delays are estimated for each frequency such that DOAs may be estimated using a least-squares technique [2]. A weighted average of these estimates is computed with the DFT magnitudes as weights. This emphasizes bins that are more likely to contain salient speech energy and so provides some robustness to reverberation. The measurements in the RVM and WDS are taken to be the unit vector whose angle is the estimated DOA and the DOA itself, respectively.

B. Simulated tracking experiments

We ran Monte Carlo simulations in a 2D room simulator with walls 5 meters in length and a T_{60} reverberation time of 165 milliseconds. The state path was generated according to (3) with the position on a circle of radius 1 meter centered on the array. The array itself was centered in the room. One

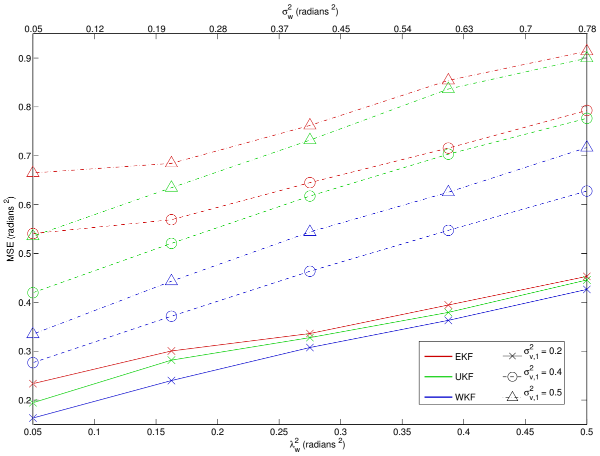

Fig. 4. MSE for EKF, UKF, and WKF over 1000 sequences for reverberant speaker tracking on the unit circle. $\sigma_{v,2}^2 = 10^{-5}$.

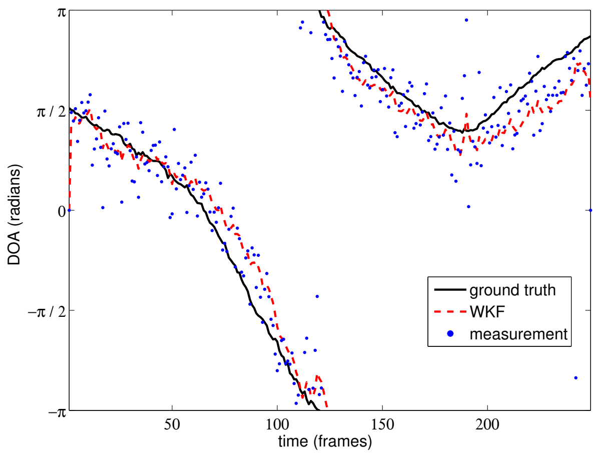

Fig. 5. WKF estimate of speaker path in real, reverberant room. $\sigma_{v,1}^2 = 5 \times 10^{-3}, \sigma_{v,2}^2 = 10^{-5}, \sigma_w^2 = 0.1$.

sentence (2-3 seconds in length) was chosen at random from the TSP speech corpus [14] for each trial.

We added noise to the estimated DOAs to (1) ensure that they conform to the filters' noise models and (2) test the filters' robustness. We attempted to match the noise levels as follows. Given a 2D Gaussian variance λ_w^2 in the RVM, the matching 1D WG variance σ_w^2 is found by fitting a WG to the angles of 10^4 samples drawn from the 2D Gaussian. The fit is reasonable but slightly overestimates the spread of the angles. Thus, our results are slightly biased in favor of the RVM.

The circular MSE for a single trial was calculated as:

$$\text{MSE} = \frac{1}{T}\sum_{t=1}^{T} \min_{l\in[-\infty,\infty]} \left(\widehat{\mu}_{t,1} - \theta_t + 2\pi l\right)^2 \quad , \qquad (20)$$

Results are summarized in Fig. 4 for various noise settings. The process and measurement noise parameters were set to the true values corresponding to either model during inference. We can observe that the WKF consistently tracks the speaker's path most accurately. This is due to the fact that the WKF infers the state path from lower-dimensional measurements.

C. Real tracking experiment

A 3-microphone array was placed on a table in the middle of a medium-sized office room with dimensions 5 × 7 × 4 meters and a T_{60} time of 25 milliseconds. 2 sentences from the TSP database were played in sequence from a loudspeaker while it was moved around the array at a radius of about 1.5 meters. High-frequency narrowband noise was simultaneously played to estimate a ground-truth DOA path. The WKF was given measurements calculated from the remaining spectrum. Fig. 5 shows the tracking results for one such experiment. As in simulations, we found that the WKF was able to track the source at least as accurately as the 2D methods.

VI. Conclusions

The wrapped Kalman filter (WKF) was introduced for tracking the DOA of a moving speaker. It was shown that the WKF is a Kalman filter whose filtered state distribution is a wrapped Gaussian. The MSE of the state path estimate is reduced by regarding the measurement as a 1D quantity embedded on the unit circle rather than as a vector in $\mathbb{R}^2$.

An approximation was used in the *correct* step to ensure that the filtered state distribution remains a WG. This was interpreted alternatively as a measurement fusion step and a good approximation to the "collapse" operation in a switching filter. Finally, we demonstrated the advantages of the WKF over conventional EKF and UKF algorithms for azimuthal speaker tracking under reverberant conditions.

References

[1] S. Gannot and T. G. Dvorkind, "Microphone array speaker localizers using spatio-temporal information," *EURASIP Journal on Advances in Signal Processing*, 2006.

[2] K. M. Varma, "Time delay estimate based direction of arrival estimation for speech in reverberant environments," M.S. thesis, Virginia Polytechnic Institute and State University, 2002.

[3] J. Benesty, J. Chen, and Y. Huang, *Topics in Signal Processing: Microphone Array Signal Processing*, vol. 1, Springer, 2008.

[4] R. E. Kalman, "A new approach to linear filtering and prediction problems," *Journal of Basic Engineering*, vol. 82, no. 1, pp. 35–45, 1960.

[5] G. Welch and G. Bishop, "An introduction to the Kalman filter," Tech. Rep., University of North Carolina at Chapel Hill, 2006.

[6] X. Xie and Y. Pi, "Phase noise filtering and phase unwrapping method based on unscented Kalman filter," *Journal of Systems Engineering and Electronics*, vol. 22, no. 3, pp. 365–372, 2011.

[7] H. Nies, O. Loffeld, and R. Wang, "Phase unwrapping using 2D-Kalman filter - potential and limitations," *IEEE International Geoscience and Remote Sensing Symposium, IGARSS 2008*, vol. 4, pp. 1213–1216, 2008.

[8] K. Mardia and P. Jupp, *Directional Statistics*, Wiley, 1999.

[9] P. Smaragdis and P. Boufounos, "Learning source trajectories using wrapped-phase hidden Markov models," *IEEE Workshop on Applications of Signal Processing to Audio and Acoustics*, pp. 114–117, 2005.

[10] A. Doucet, N. de Freitas, K. Murphy, and S. Russell, "Rao-Blackwellised particle filtering for dynamic bayesian networks," *Uncertainty in AI*, 2000.

[11] J. Traa, "Multichannel source separation and tracking with phase differences by random sample consensus," M.S. thesis, University of Illinois at Urbana-Champaign, 2013.

[12] Q. Gan and C. Harris, "Comparison of two measurement fusion methods for Kalman-filter-based multisensor data fusion," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 37, no. 1, pp. 273–279, 2001.

[13] K. P. Murphy, "Switching Kalman filters," Tech. Rep., University of British Columbia, 1998.

[14] P. Kabal, "TSP speech database," 2002, Telecommunications and Signal Processing Lab, McGill University.