Tampere University

Mikko Lager

AUDIO SOURCE POSITIONING BASED ON ANGLE OF ARRIVAL MEASUREMENTS

Faculty of Information Technology and Communication Sciences
Master of Science Thesis
February 2020

ABSTRACT

Mikko Lager: Audio Source Positioning Based on Angle of Arrival Measurements
Master of Science Thesis
Tampere University
Degree Programme in Science and Engineering
February 2020

Estimating position is done in various contexts from locating phones with GPS to locating boats using hydrophones. In this thesis we study estimating audio source position based on angle of arrival measurements. Multiple different filters can be used on measured angles of arrival to deduce the position of the source. The filter to be used in this work was chosen to be the particle filter. Even though particle filter is computationally more heavy than many other filters, modern computers can simulate hundreds of particles in a short time without too much of an effort. We introduce the reader to the use of particle filter in positioning, along with theoretical background of it and positioning in a more general sense.

The data in this work is recorded in either an anechoic chamber or a room that has no special equipment installed to enhance audio quality in it. The measurements are done with a mobile device with four microphones. Audio source in the anechoic chamber is a loudspeaker playing speech or a person speaking and walking randomly in the room. If the data contains noise, it is played from loudspeakers in the same space as the source is located in. Another type of data handled in this work is measured outside in a racing event where multiple cars passed the measurement device as well as generated data with multiple sources.

The data is handled as a mixture between von Mises and uniform distribution. An important parameter of von Mises distribution is a variable called κ, which tells the concentration of the distribution. In this work we show and prove a way to estimate said variable with maximum likelihood method. Additionally, we introduce the reader to mathematical background of particle filter and positioning in more general sense. Results given by the particle filter depend on the chosen value of κ along with chosen q-value, which tells the smoothness of the result, and measurement model. Finally, we present and compare the results obtained by constant velocity and random walk models with several different q-values.

Keywords: Particle filter, angle of arrival measurement, position estimation, von Mises distribution

TIIVISTELMÄ

Mikko Lager: Audiolähteen paikannus saapumiskulmamittausten avulla
Diplomityö
Tampereen yliopisto
Teknis-luonnontieteellinen koulutusohjelma
Helmikuu 2020

Paikan estimointia tehdään useassa eri yhteydessä puhelimen paikannuksesta GPS:n avulla aina laivojen paikannukseen hydrofoneja käyttämällä. Tässä diplomityössä tutkitaan audiolähteen paikannusta saapumiskulmamittausten avulla. Saapumiskulmamittauksiin voidaan käyttää useita erilaisia suodattimia audiolähteen paikan selvittämiseksi. Tähän työhön käytettäväksi suodattimeksi valittiin partikkelisuodatin. Vaikka partikkelisuodatin vaatiikiin enemmän laskentatehoa kuin monet muista suodattimista, nykykoneet voivat vaivattomasti simuloida satoja partikkeleita lyhyessä ajassa. Lukija tutustutetaan partikkelisuodattimen käyttöön paikannuksessa, sekä kyseisen suodattimen ja paikannuksen teoriaan yleisemmällä tasolla.

Työssä käytettävä data on äänitetty joko kaiuttomassa kammiossa tai huoneessa, jota ei ole varustettu erilaisilla äänenlaatua parantavilla ominaisuuksilla. Kaikki mittaukset on tehty mobiililaitteella, jossa on neljä mikrofonia. Audiolähde kaiuttomassa tilassa on kaiutin, josta soitetaan puhetta ja huoneessa ihminen, joka puhuu ja kävelee satunnaisesti ympäri huonetta. Mikäli datassa on kohinaa, se on soitettu samassa tilassa olevista kaiuttimista. Tämän datan lisäksi työssä käsitellään dataa, joka on mitattu autokisassa, jossa mittauslaitteen ohi ajoi usea auto peräjälkeen, sekä generoitua dataa joka sisältää useamman äänilähteen.

Mitattua dataa käsitellään von Mises jakauman ja tasajakauman välisenä yhdistettynä jakaumana. Tärkeä osa von Mises jakaumaa on parametri κ, joka kertoo jakauman keskittämisen. Tässä työssä näytetään ja todistetaan tapa approksimoida tätä parametria käyttämällä suurimman uskottavuuden menetelmää. Lisäksi lukijalle tulee tutuksi partikkelisuodattimen sekä paikannuksen matemaattinen tausta. Partikkelisuodattimen antamat tulokset riippuvat valitun parametrin κ lisäksi valitusta q-arvosta, joka vaikuttaa tulosten tasaisuuteen, ja valitusta mittausmallista. Lopuksi esitellään ja vertaillaan vakionopeusmallin ja pelkkään paikkaan sidotun mallin antamia tuloksia eri q-arvoilla.

Avainsanat: Partikkelisuodatin, saapumiskulmamittaukset, paikan estimointi, von Mises-jakauma

PREFACE

First of all I want to thank everyone involved in the making of this thesis. Thank you for your contributions, thank you for the support and thank you for the advice.

Second of all I want to thank everyone involved in my life these past years of my studies. Without you I probably would have graduated couple years earlier and now be miserable in a job I don't enjoy, with no friends and all this experience. You have made me who I am today, and made me appreciate the things worth of appreciation. Life is more than just work.

I also own my gratitude to my fiancée Anna, without whom I would never had finished this thesis. I love you, and thank you for keeping me from going insane.

Special thanks are also to be given to a chat group Fa across Fa account xAQUUAC, from which I've gotten various advice in various fields of life and countless of moments of joy. This is a safe space to laugh...and it makes one laugh.

Tampere, 12th February 2020

Mikko Lager

CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF SYMBOLS AND ABBREVIATIONS

ϕ_k	mean direction angle at time k
π	importance distribution
$C(\kappa)$	normalization constant in von Mises-Fisher distribution
Δt	step length $t_k - t_{k-1}$
$\delta(\cdot)$	Dirac delta function
E	expected value
F	state transition matrix
$f(x \mid y)$	conditional probability density function
$f(x)$,$g(x)$	probability density function of random variable x
$I_\alpha(\cdot)$	modified Bessel function of the first kind of order α
$I_{n\times n}$	n times n identity matrix
$\int_A$	integral over the set A
κ	concentration parameter in von Mises-Fisher distribution
MATLAB	matrix laboratory, a computing environment
ML(E)	Maximum likelihood (estimation)
$\boldsymbol{\mu}$, μ	mean direction vector, mean direction value
$\mathbb{N}$	set of natural numbers
$\mathcal{N}$	normal distribution
n, N	natural number
P	probability
$p(x \mid y)$	conditional probability density function
p	2-dimensional vector with azimuth and elevation angles in it
ϕ, θ	angles
Q_k	covariance matrix
Q_c	diffusion matrix
$\mathbb{R}$	set of real numbers
R_k	component of measurement in the known direction
$\boldsymbol{\Sigma}$	covariance matrix
σ	variance

σ_{azi},σ_{ele}	variance of measured angles
$\theta_{1:k}$	measured angles
$()^T$	transpose
$\mathcal{U}$	uniform distribution
v	2-dimensional vector with angle velocities in it
VMF	von Mises-Fisher distribution
$w^{(i)}$	normalized weight of particle i
$x \in A$	x is in set A
$x^{(i)}$	random sample i, generated by a probability density function
$x_{1:k}$	a set of vectors $\{x_1, \ldots, x_k\}$
$y_{azi,k}$, $y_{ele,k}$	measurement vector of azimuth and elevation angles at time k

1 INTRODUCTION

Estimating the direction of an audio source is a process where the sound is used to determine the location of the source [28]. Tracking and estimating the direction of an audio source is something that people do daily: following a speech of another person, knowing the rough position of a nearby car passing by without actually seeing it, or noticing where a coin dropped on the floor. For humans this skill is based on binaural hearing and mainly the time difference of arriving sound waves [19]. While human physiology makes the estimating the position possible within those skills limits, computers provide a variety of other possibilities to estimate position of an audio source, as several parameters can be measured at once and with multiple microphones. These measurements can be of form of time of arrival, time difference of arrival, angle of arrival or received signal strength (pressure in case of audio waves), for example [29].

The motivation behind tracking an audio source, besides of it being an interesting field of research, is mainly the possibilities it gives us. Some of the previous research on this subject include path planning for automated robots and estimating positions of passing vehicles in ocean environments [11, 16]. Tracking an audio source can also be useful in speech recognition or audio source separation, which in turn are used in, for instance, automatic home assistants [3]. In this work we focus on the mathematical principles of particle filter and the use of said filter in positioning audio sources based on angle of arrival measurements.

In Chapter 2 we look into the theoretical background of the problem and introduce the used algorithms. We also introduce the assumed distributions of measurements and deduce the maximum likelihood estimation of the concentration parameter in von Mises-Fisher distribution. On top of this the principles of how particle filter works are shown. Chapter 3 introduces the reader to state system and observation models in previous research done on similar subjects, as well as the models used in this work, which are later on used in a particle filter.

Chapter 4 is divided in three parts, first of which includes the introduction to used data; where and in what kind of conditions the measurements are done. In the second part we focus on the results given by maximum likelihood method on position estimation and κ approximation. The third part includes the results given by particle filter and comparison between used values and models. In the last part we also discuss the possibilities of tracking multiple targets with particle filters.

The algorithms used in this work have been created using MATLAB and tested with both measured and generated data.

2 THEORY

For linear and non-linear measurements different models and filters are used to get an accurate approximation of the state of a time-varying system. In this chapter we will go through the theory of the state space and measurement models as well as filters.

2.1 General problem

In positioning we have multiple measurements which are used to calculate the estimate of the source position. In addition to the position estimate the accuracy and error are assessed. The state and measurements of the system are described by models, which are formed with the following definitions.

Definition 2.1 ([33, pp. 238-239])**.** A random variable $x \in \mathbb{R}^n$ is *continuous* if there exists a non-negative probability density function $f_x(x) \geq 0$ so that

$$\int f_x(x)\, dx = 1, \tag{2.1}$$

and the probability for an event $x \in A$ can be calculated with

$$P\{x \in A\} = \int_A f_x(x)\, dx. \tag{2.2}$$

Let's write the time series as a vector $\{x_1, x_2, \ldots\}$ and the noisy measurement as $\{y_1, y_2, \ldots\}$. Based on observed measurements we can estimate the unknown state of the system. A stochastic process is a series of random variables $x_k \in \mathbb{R}^n$, where $k \geq 1$ tells us the time t_k. In a continuous process $k \in [1, \infty)$ and in a discrete process $k \in \{1, 2, \ldots\}$. A stochastic process is called Markov process, if the probability of an event depends only of the state reached in previous event. [7, pp. 23-24] Next we will define Markov process mathematically.

Definition 2.2 ([17, pp. 30-31])**.** The *conditional probability density function* of x given y is formed by

$$f(x \mid y) = \frac{f(x, y)}{f(y)}. \tag{2.3}$$

We can calculate conditional probabilities by inserting this into Equation (2.2):

$$P\{x \in A \mid y\} = \int_A f(x \mid y)\, dx. \tag{2.4}$$

Definition 2.3. [20, p. 25] A stochastic process is called a *Markov process* if

$$f(x_k \mid x_1, \ldots, x_{k-1}) = f(x_k \mid x_{k-1}). \tag{2.5}$$

With these definitions we can formulate the problem. The state and measurements are modelled as stochastic processes $\{x_k, k \in \mathbb{N}\}$ and $\{y_{k+1}, k \in \mathbb{N}\}$ with precondition x_0. The state is formed as a difference equation

$$x_{k+1} = f_k(x_k, q_k), \tag{2.6}$$

where f_k is a continuously differentiable function and q_k the system noise. Similarly the measurement is formed as

$$y_k = h_k(x_k, r_k), \tag{2.7}$$

where h_k is continuously differentiable and r_k is the measurement noise. [2]

Optimal filtering is an inversion problem where we don't know the time series $\{x_1, x_2, \ldots\}$ which is observed through noisy measurements $\{y_1, y_2, \ldots\}$. The object of statistical inversion is to approximate the hidden state x_t based on measurements $y_{1:k} = \{y_1, \ldots, y_k\}$. [27, pp. 8-9] Filtering utilizes measurements up until and including present time to calculate the approximation. It is done in two steps, called prediction and update steps. In prediction step we have to solve the distribution

$$f(x_k \mid y_{1:k-1}) \tag{2.8}$$

based on measurements $y_{1:k-1}$. In update step measurements $y_{1:k}$ are observed, the problem is to solve the filtering, or the marginal posterior, distribution

$$f(x_k \mid y_{1:k}). \tag{2.9}$$

The initial distribution is independent of the measurements, as they haven't been made yet.[27, p. 54]

To find the conditional probability density functions we use Bayes' rule [21, pp. 175-176] with which we can compute the update of the state (2.9) by

$$f(x_k \mid y_{1:k}) = \frac{f(y_k \mid x_k) f(x_k \mid y_{1:k-1})}{\int f(y_k \mid x_k) f(x_k \mid y_{1:k-1}) dx_k}, \tag{2.10}$$

where $f(y_k \mid x_k)$ is the likelihood function defined by the measurement model equation shown in Equation (2.7)[25, p. 5]. Given the dynamic model, the function (2.9) satisfies a recursion, where the state x_k has a predicted distribution based on Chapman-Kolmogorov

equation [10, p. 6]

$$f(x_k \mid y_{1:k-1}) = \int f(x_k \mid x_{k-1}) f(x_{k-1} \mid y_{1:k-1}) dx_{k-1}. \tag{2.11}$$

In the above Equation (2.11) $f(x_k \mid x_{k-1})$ is defined by the system model Equation (2.6).[25, p. 5]

2.2 Distributions

Usually the noise of measurements follow some distribution that can be formulated mathematically. In this work the measurements are assumed to have von Mises-Fisher distribution combined with a uniform distribution. In addition to these two distributions we need to introduce the normal distribution, which is used in the state system model.

Definition 2.4 ([12, p. 43])**.** If the probability density function of a random variable $\theta \in (-\pi, \pi]$ is

$$f(\theta) = \frac{1}{2\pi}, \tag{2.12}$$

it is said to be *uniformly distributed* on a circle. If the random variable is uniformly distributed on a circle, it is denoted as $\theta \sim \mathcal{U}((-\pi, \pi])$.

Definition 2.5 ([17, p. 32],[22, p. 231])**.** A random variable $x \in \mathbb{R}^n$, $n \in \mathbb{N} \setminus \{0\}$, with expected value $\mathsf{E}(x) = \mu \in \mathbb{R}^n$ and a symmetric positive definite covariance matrix $V(x) = \mathbf{\Sigma} \in \mathbb{R}^{n \times n}$ is said to have a *normal*, or *Gaussian, distribution* if its probability density function is

$$f(x) = \frac{1}{\sqrt{(2\pi)^n \det(\mathbf{\Sigma})}} e^{-\frac{1}{2}(x-\mu)^{\mathsf{T}} \mathbf{\Sigma}^{-1} (x-\mu)}. \tag{2.13}$$

In this case we write

$$x \sim \mathcal{N}(\mu, \mathbf{\Sigma}). \tag{2.14}$$

Definition 2.6 ([4])**.** A unit vector $x \in \mathbb{R}^n$, $n \geq 2, n \in \mathbb{N}$, is said to be *n-variate von Mises−Fisher (VMF) distributed* if its probability density function is

$$f(x) = C(\kappa) e^{\kappa \left(\mu^T x\right)}, \tag{2.15}$$

where μ is the mean direction as a unit vector,

$$C(\kappa) = \frac{\kappa^{n/2-1}}{(2\pi)^{n/2} I_{n/2-1}(\kappa)} \tag{2.16}$$

is a normalization constant, $\kappa \geq 0$ a concentration parameter and $I_k(\cdot)$ is modified Bessel function of the first kind of order k. VMF distribution is denoted as $x \sim \text{VMF}(\mu, \kappa)$.

If $n = 2$, we have a one-dimensional distribution on a circle and the Equation (2.15)

reduces to

$$f(x) = \frac{e^{\kappa(\mu^T x)}}{2\pi I_0(\kappa)}, \tag{2.17}$$

as can be seen from [6] and is said to have von Mises or *circular normal* distribution [13]. In this case the distribution describes only one of the angles at a time, and the estimation will be done separately on azimuth and elevation angles. Also it's worth to note that κ gets a different value depending on which angle is used. Another form of the Equation (2.17) using angles instead of vectors is obtained by using dot product of vectors and the fact that the vectors are unit vectors. This form is

$$f(\phi) = \frac{e^{\kappa \cos(\phi - \phi_0)}}{2\pi I_0(\kappa)}, \tag{2.18}$$

where ϕ is the measured mean direction angle and $\phi_0 \in (0, 2\pi]$ is the known direction.

In the case of a sphere, where $n = 3$, Equation (2.16) becomes [24, p. 229]

$$C(\kappa) = \frac{\kappa}{4\pi \sinh(\kappa)}. \tag{2.19}$$

Intuitively the three dimensional von Mises-Fisher distribution is easy to understand. The mean direction can be expressed as two normally distributed angles with a certain connection between κ and variance (σ^2) of the angles. This connection is

$$\kappa = \frac{1}{\sigma^2}, \tag{2.20}$$

with the assumption that both angles are similarly distributed. The derivation of this connection can be found on [30, pp. 10-12].

2.3 Maximum likelihood estimation and particle filter

The aim is to estimate the measured system state while paying attention to the noise of the system. This can be achieved by using filters and estimators, which calculate the directions as expected values. In this section we will go through the particle filter, maximum likelihood estimation and how to estimate κ of von Mises-Fisher distribution with maximum likelihood method.

2.3.1 Maximum likelihood estimation and the estimation of κ

The maximum likelihood estimation (MLE) is used to find out the unknown parameters of a probability model. In this work it is used to find the parameters of the Gaussian and von Mises-Fisher distribution models, namely variance (Σ) and mean value (μ) in Equation (2.13) and mean direction vector ($\boldsymbol{\mu}$) and kappa (κ) in Equation (2.15) based on given data. Note that in this case $\boldsymbol{\Sigma} = \sigma \in \mathbb{R}^{1\times 1}$, as maximum likelihood estimation is used for

static positioning i.e. for only one time step at once, without the previous measurements affecting the present.

MLE is based on the Maximum Likelihood Principle: given the sample x and a model $f(x;\theta)$, choose the estimator of θ so that it maximizes the likelihood function. By definition the MLE of θ is a solution to the maximization problem $\max_{\theta}[L(\theta)]$, where L is the likelihood function.[26, p. 226] Next we will show how to apply MLE in estimation of κ parameter in Equation (2.15).

With a prior distribution, given by a conjugate to Equation (2.18),

$$f\left(\phi,\kappa\right) \propto I_0\left(\kappa\right)^{-c} e^{\kappa R_0 \cos(\phi-\phi_0)}, \tag{2.21}$$

where ϕ is the mean direction angle, $R_0 \geq 0$ is the component of the measurement in the known direction, $\phi_0 \in (0, 2\pi]$ and $c \geq 0$ we get the posterior distribution proportional to

$$f\left(\phi,\kappa|\theta_{1:k}\right) \propto \frac{1}{I_0(\kappa)^{c+k}} e^{\kappa R_k \cos(\phi-\phi_k)}, \tag{2.22}$$

given the conditionally independent angle measurements $\theta_{1:k}$, given ϕ_0, κ. Only the direction is examined earlier in the Equation (2.18), and thus $R_0 = 1$ is not shown in said equation. However, for the sake of generality, it is considered here. In the above Equation (2.22) ϕ_k and R_k are obtained from

$$\begin{aligned} R_k \cos(\phi_k) &= R_0 \cos(\phi_0) + \sum_{i=1}^{k} \cos(\theta_i), \text{ and} \\ R_k \sin(\phi_k) &= R_0 \sin(\phi_0) + \sum_{i=1}^{k} \sin(\theta_i), \end{aligned} \tag{2.23}$$

and $\theta_{1:k}$ are the measured angles.[15] Now, using these Equations (2.22)-(2.23) we can prove the following theorem.

Theorem 2.1. *Maximum a posteriori estimation of*

$$\kappa = A^{-1}\left(\frac{R_k}{c+k}\right), \tag{2.24}$$

where $A = \frac{I_1(\kappa)}{I_0(\kappa)}$ *and* $c \geq 0$.

Proof. We get the maximum a posteriori estimation for the concentration parameter κ from the modal point of posterior distribution shown in Equation (2.22), and is obtained by solving

$$\frac{d}{d\kappa} f(\phi,\kappa|\theta_{1:k}) = 0. \tag{2.25}$$

In general for non-zero functions we can write

$$f'(x) = 0 \Leftrightarrow 0 = \frac{f'(x)}{f(x)} = \frac{d}{dx}\log(f(x)). \tag{2.26}$$

Thus, as posterior distribution in Equation (2.22) is greater than zero, using the identity shown in Equation (2.26) we get the same result by the logarithmic derivative

$$\frac{d}{d\kappa}\log(f(\phi,\kappa|\theta_{1:k})) = \frac{d}{d\kappa}\log\left[\frac{e^{\kappa R_k}}{I_0(\kappa)^{c+k}}\right] = 0, \tag{2.27}$$

which simplifies the future calculations. Note that compared to Equation (2.22) the cosine in the exponent is left out; this is because the angles represent the average direction and we get no loss of generality assuming the direction to be zero [15]. From Equation (2.27) we get

$$\frac{d}{d\kappa}\left[\log(e^{\kappa R_k}) - \log(I_0(\kappa)^{c+k})\right] = 0, \tag{2.28}$$

which further can be developed into

$$\frac{d}{d\kappa}\log(e^{\kappa R_k}) = \frac{d}{d\kappa}\log(I_0(\kappa)^{c+k}). \tag{2.29}$$

This gives us

$$R_k = (c+k)\frac{\frac{d}{d\kappa}I_0(\kappa)}{I_0(\kappa)}, \tag{2.30}$$

and by using the identity $\frac{d}{d\kappa}I_0(\kappa) = I_1(\kappa)$, which can be found in [1, p. 376], we get

$$\frac{R_k}{c+k} = \frac{I_1(\kappa)}{I_0(\kappa)}. \tag{2.31}$$

By using the notation $\frac{I_1(\kappa)}{I_0(\kappa)} = A(\kappa)$, we have

$$\frac{R_k}{c+k} = A(\kappa), \tag{2.32}$$

from where we get κ as

$$\kappa = A^{-1}\left(\frac{R_k}{c+k}\right). \tag{2.33}$$

To make sure that this point is a maximum instead of a minimum or a saddle point, we have to take a look into the second derivative of the function. Similarly to what we did earlier, we get the first derivative by

$$\frac{d^2}{d\kappa^2}\log\left[\frac{e^{\kappa R_k}}{I_0(\kappa)^{c+k}}\right] = \frac{d}{d\kappa}\left[R_k - (c+k)\frac{I_1(\kappa)}{I_0(\kappa)}\right], \tag{2.34}$$

from where we can continue to

$$\frac{d}{d\kappa}\left[R_k - (c+k)\frac{I_1(\kappa)}{I_0(\kappa)}\right] = -(c+k)\left[\frac{I_0(\kappa) - \frac{1}{\kappa}I_1(\kappa)}{I_0(\kappa)} - \frac{I_1(\kappa)^2}{I_0(\kappa)^2}\right]. \tag{2.35}$$

In the above Equation (2.35) we can use the same notation as earlier and replace $\frac{I_1(\kappa)}{I_0(\kappa)}$

with $A(\kappa)$, giving us

$$\begin{aligned} -(c+k)\left[\frac{I_0(\kappa)-\frac{1}{\kappa}I_1(\kappa)}{I_0(\kappa)}-\frac{I_1(\kappa)^2}{I_0(\kappa)^2}\right] &= -(c+k)\left[\frac{I_0(\kappa)}{I_0(\kappa)}-\frac{\frac{1}{\kappa}I_1(\kappa)}{I_0(\kappa)}-\left(\frac{I_1(\kappa)}{I_0(\kappa)}\right)^2\right] \\ &= -(c+k)\left[1-\frac{I_1(\kappa)}{\kappa I_0(\kappa)}-\left(\frac{I_1(\kappa)}{I_0(\kappa)}\right)^2\right] \\ &= -(c+k)\left[1-\frac{1}{\kappa}A(\kappa)-A^2(\kappa)\right]. \end{aligned} \tag{2.36}$$

Now, as $A(\kappa)$ is a monotonically increasing function from 0 to 1 as κ goes from 0 to ∞, and $\lim_{\kappa\to 0}\frac{1}{\kappa}A(\kappa)=\frac{1}{2}$, the values of the bracketed term in Equation (2.36) is limited in $\left(0,\frac{1}{2}\right)$ [5]. As $c+k>0$ always with the earlier set limitations as well, we can safely say that the second derivative is negative with every $\kappa>0$, especially at the location given by Equation (2.33), thus making the modal point a maximum [9, pp. 158-159]. □

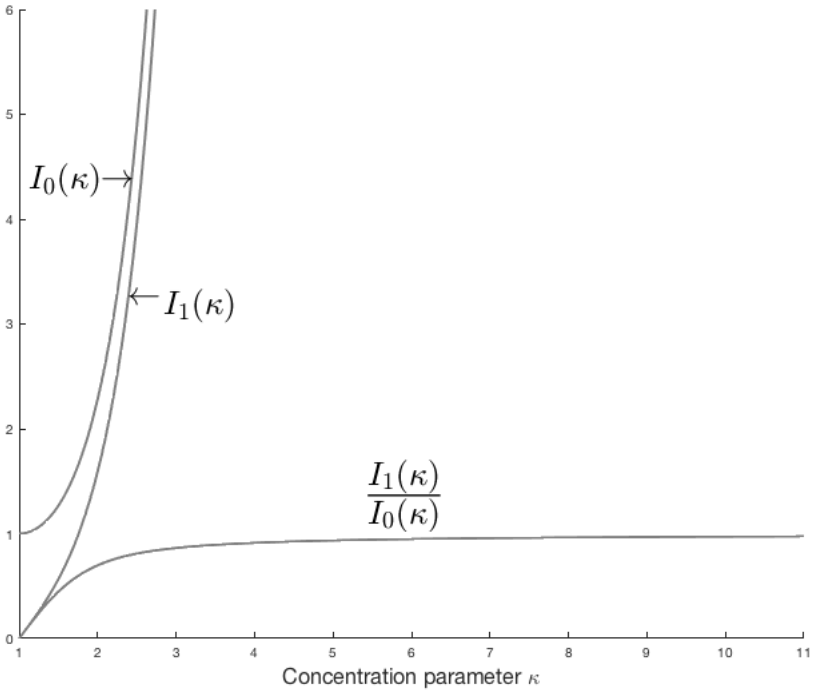

Figure 2.1. *Behavior of $I_0(\kappa)$, $I_1(\kappa)$, and their quotient*

The monotonicity of $A(\kappa)$ is demonstrated in Figure 2.1, along with the behavior of modified Bessel functions of the first kind of order 0 and 1. The estimation of κ given by Equation (2.33) is used throughout this work. In this work the inverse function of A is computed by using a function that finds out the root of a nonlinear function in an interval. The upper limit of said interval is chosen so that the true κ is certainly below it. So in short we simply solve $A(\kappa)-\frac{R_k}{c+k}=0$ for κ.

2.3.2 Particle filter

Particle filter is based on a Monte Carlo simulation where posterior distribution is estimated with particles. These (state)particles include the vector and the weight of the state. In a Monte Carlo simulation we have N independent random samples

$$x_k^{(i)} \sim f\left(x_k \mid y_k\right), \tag{2.37}$$

where $i = 1, \ldots, N$. The expectation

$$\mathsf{E}\left[g\left(x_k\right) \mid y_k\right] = \int g\left(x_k\right) f\left(x_k \mid y_k\right) dx, \tag{2.38}$$

where g is an arbitrary integrable function and $f\left(x_k \mid y_k\right)$ is the posterior distribution based on measurements y_k, can now be estimated as

$$\mathsf{E}\left[g\left(x_k\right) \mid y_k\right] \approx \frac{1}{N} \sum_{i=1}^{N} g(x_k^{(i)}). \tag{2.39}$$

[27, pp. 116-117] With known measurement model $f\left(y_k \mid x_k\right)$ and prior distribution $f\left(x_k \mid x_{k-1}\right)$ we can form the estimation of posterior distribution by using importance sampling. We approximate the distribution with importance distribution $\pi\left(x_k \mid y_k\right)$, from which we can draw samples

$$x_k^{(i)} \sim \pi\left(x_k \mid y_k\right), \tag{2.40}$$

where $i = 1, \ldots, N$. With the importance distribution we can rewrite the Equation (2.38):

$$\begin{aligned} \mathsf{E}\left[g\left(x_k\right) \mid y_k\right] &= \int g\left(x_k\right) f\left(x_k \mid y_k\right) dx \\ &= \int \left[g\left(x_k\right) \frac{f\left(x_k \mid y_k\right)}{\pi\left(x_k \mid y_k\right)}\right] \pi\left(x_k \mid y_k\right) dx. \end{aligned} \tag{2.41}$$

The Equation (2.41) is now the expectation of the bracketed term over the distribution $\pi\left(x_k \mid y_k\right)$, which allows us to form a Monte Carlo estimation to it. After generating the samples we calculate the weights as

$$w^{*(i)} = \frac{f\left(y_k \mid x_k^{(i)}\right)}{\pi\left(x_k^{(i)} \mid y_k\right)} f(x_k^{(i)} \mid y_{k-1}), \tag{2.42}$$

which are then normalized, resulting in

$$w^{(i)} = \frac{w^{*(i)}}{\sum_{j=1}^{N} w^{*(j)}}. \tag{2.43}$$

The posterior expectation can now be formed with these normalized weights as

$$\mathrm{E}\left[g\left(x_k\right) \mid y_k\right] \approx \sum_{i=1}^{N} w^{(i)} g\left(x_k^{(i)}\right) \tag{2.44}$$

and the approximation of probability density function of the posterior distribution

$$p\left(x_k \mid y_k\right) \approx \sum_{i=1}^{N} w^{(i)} \delta(x_k - x_k^{(i)}), \tag{2.45}$$

where δ is the Dirac delta function. [27, pp. 118-120]

With sampling we can sometimes meet a *degeneracy problem*, in which all the particles have near-zero or zero weights. This can be solved with resampling. In resampling the algorithm eliminates the low importance particles and multiplies the high importance ones. Formally we replace Equation (2.45) with

$$p\left(x_k \mid y_k\right) = \sum_{i=1}^{N} \frac{n_i}{N} \delta(x_k - x^{(i)}), \tag{2.46}$$

where n_i is the number of copies of the particle $x^{(i)}$ in the new set of particles. There is, of course, multiple ways to execute the resampling and the most frequently encountered methods include stratified, multinomial, systematic and residual resampling.[18] In this work we will use multinomial resampling to reweight the particles.

In multinomial resampling the main idea is to generate N independent random numbers $u^{(n)}$, where $n \in \{1, \ldots, N\}$, from the uniform distribution $(0, 1]$ and use them to select the particles from the posterior distribution represented by Equation (2.45). The particle $x^{(m)}$ is chosen, when it satisfies the condition

$$\sum_{k=1}^{m-1} w^{(k)} \leq u^{(n)} < \sum_{k=1}^{m} w^{(k)} \tag{2.47}$$

so that the probability of selecting $x^{(m)}$ is equal to $u^{(n)}$ being in the interval limited by the cumulative sum of the normalized weights, thus lowering the number of close to zero or zero weight particles in the process. [18, 23]

In some cases resampling is done only when the number of low weight particles rise above a certain treshold, but in this work, however, resampling is done on every step of the particle filter. The reasoning behind this is that the number of generated particles is low and the computation is not too expensive.

A basic particle filter, so called bootstrap filter, of N particles can be written as:

Algorithm 1 Particle Filter

procedure PF
  Generate N particles $x_0^{(i)}$ from the prior distribution $x \sim f(x)$
  Set weight $w_0^{(i)} = \frac{1}{N}$ for every $i = 1, \ldots, N$
  for $k = 1 : T$ **do**
    Generate particles $x_k^{(i)}$ from the state model distribution $f(x_k|x_{k-1}^{(i)})$
    Calculate new weights based on measurement and normalize them as shown in Equations (2.42) and (2.43)
    Resample
    Calculate expected value of posterior distribution x_{exp} with Equation (2.44)
  end for
end procedure

The x_{exp} given by Algorithm 1 is the estimated direction. The importance distribution, where the weights are computed from, in this work is a mixture distribution of von Mises and uniform distributions. The particle filter used in this work can be found in Appendix B implemented in MATLAB.

3 USED MODELS IN ACOUSTIC TRACKING

In this chapter we introduce the reader to the state system model and measurement models used in tracking an audio source. First we show models in more general sense in what has been done before, and afterwards present the models used in this work. In this work we test two different state system models: random walk and constant velocity models.

3.1 Literature overview of used models

Previous work shows that audio source positioning has been done for both circular and spherical data, as is case in this work, as well as for more Euclidean two- or three-dimensional spaces. The main difference between different geometries lies in the distributions used in the system and measurement models. In spherical and circular cases most common distributions are von Mises(-Fisher) and wrapped Gaussian, or some mixture between these, while in other cases the distribution generally is Gaussian or some variation of it. [31, 32, 34]

The general state system and measurement models were introduced in Chapter 2 as Equations (2.6) and (2.7). Depending on the problem these equations can be different. The differences come mainly from distribution of noise and movement, but also from how the position is shown. Some of the previous work done on audio source tracking use azimuth and elevation angles to show position, while others use a position vector. For example, in [31] Traa and Smaragdis used the angle representation while using wrapped Kalman filter for azimuthal speaker tracking, and in [32] they used a position vector to show position while tracking multiple speakers with factorial von Mises-Fisher filter. In both works the used distributions were the von Mises-Fisher and Gaussian distribution. The information about position within both angles and the vector is the same, but the mathematical background changes as same formulas don't work in both (take Equation (2.17) vs. Equation (2.18), for example). In this work we use azimuth and elevation angles to show the position of the audio source.

In addition to how the information is presented and used distributions the models differ in how the target is assumed to move. Models can also take into account velocity, or even the change of velocity if need be. For instance, in [8] Cevher et al. show and use a more complex model with velocity taken into account among other things when tracking

a target using joint acoustic video system. In this work we present two different state system models, one of which assumes the target to be in near-constant motion and the other assumes a random walk motion.

3.2 Models used in this work

In this section we present the models used in this work. First we show both used state system models, and secondly the used measurement model with the mixture distribution between von Mises and spherical uniform distributions.

3.2.1 State system models

A model that represents the system state is dependent on the system that is under observation. In this thesis we focus on two kinds of models. One of which includes both the location and the (near-constant) velocity of the target and the other that is based only on the location of the target, without taking velocity into account.

In the random walk model we have the location of the source

$$x = \begin{bmatrix} y_{azi,k} \\ y_{ele,k} \end{bmatrix}, \tag{3.1}$$

represented by the azimuth and elevation angles. The model itself is in the form of

$$x_k = F_{k-1}x_{k-1} + q_{k-1}, \tag{3.2}$$

where

$$F_{k-1} = \begin{bmatrix} 1 & \Delta t \\ 0 & 1 \end{bmatrix} \tag{3.3}$$

is the state transition matrix with a step length $\Delta t = t_k - t_{k-1}$ and $q_{k-1} \sim N(0, Q_{k-1})$ is a normally distributed error. The covariance matrix Q_k for this model is a diagonal matrix with variances of the angles on the diagonal

$$Q_k = \begin{bmatrix} \sigma^2_{azi} & 0 \\ 0 & \sigma^2_{ele} \end{bmatrix}. \tag{3.4}$$

By using the same notation as shown in Algorithm 1, we can write the model in Equation (3.2) in form

$$f(x_k \mid x_{k-1}) = x_{k-1} + \Delta t q_{k-1}. \tag{3.5}$$

In the constant velocity model we use a tuple, which is of the form

$$x = \begin{bmatrix} p \\ v \end{bmatrix}, \tag{3.6}$$

where p is the 2-dimensional vector of the azimuth and elevation angles and v the 2-dimensional angular velocity which contains the velocity of both angles separately. Now the state system model itself is given by

$$x_k = F_{k-1}x_{k-1} + q_{k-1}. \tag{3.7}$$

In the model the state transition matrix F_{k-1} is

$$F_{k-1} = \begin{bmatrix} I_{2\times 2} & \Delta t I_{2\times 2} \\ 0 & I_{2\times 2} \end{bmatrix}, \tag{3.8}$$

with step length of $\Delta t = t_k - t_{k-1}$ and q_{k-1} a normally distributed error $q_{k-1} \sim N(0, Q_{k-1})$. [2, pp. 43-44]

The covariance matrix Q_{k-1} of the error is calculated with

$$Q_{k-1} = \begin{bmatrix} \frac{\Delta t^3}{3}Q_c & \frac{\Delta t^2}{2}Q_c \\ \frac{\Delta t^2}{2}Q_c & \Delta t Q_c \end{bmatrix}, \tag{3.9}$$

where the diffusion matrix Q_c is same as the diagonal matrix shown in Equation (3.4). [14]

Similarly as with the random walk model, we can write the Equation (3.7) by using notation

$$f(x_k \mid x_{k-1}) = x_{k-1} + \Delta t v_{k-1}, \tag{3.10}$$

where v_{k-1} is the velocity calculated during the previous step. In addition to position update, we have to adjust the velocity as well by calculating

$$f(v_k \mid v_{k-1}) = v_{k-1} + \Delta t q_{k-1}. \tag{3.11}$$

3.2.2 Measurement model

Observation or measurement model is a model which describes the dependency between measurement y_k and current state x_k [27, p. 10]. In this thesis the measurement model is based on the two or three dimensional von Mises-Fisher distribution combined with a uniform distribution.

In this work the measurement vector has two variables; the azimuth and elevation angles:

$$y_k = \begin{bmatrix} y_{azi,k} \\ y_{ele,k} \end{bmatrix},$$

which are obtained by simulating data with known location of the source or from real measurements. Each angle is assumed to be von Mises distributed with uniformly distributed noise. The amount of noise depends on measurement, and is calculated as a percentage of measurements. The percentage used is obtained by comparing the amount of low-confidence measurements to all measurements thus far. Combining the uniform distribution with Equation (2.15) we get the probability density function of the measured angle as

$$f(\phi) = (1-\alpha)\frac{e^{\kappa\cos(\phi-\phi_0)}}{2\pi I_0(\kappa)} + \frac{\alpha}{2\pi}, \tag{3.12}$$

which is used to calculate the weights in particle filter.

Given the measured angle θ the measurement model is thus

$$y_{azi,k} = \theta + e_n, \tag{3.13}$$

where e_n is distributed according to the mixture distribution shown in Equation (3.12), and similarly for $y_{ele,k}$.

4 RESULTS

In this chapter we will go through the results given by the estimation algorithms for a few different datasets as well as introduce the reader to used data and measurement conditions.

4.1 Introduction to the used data

The real data used in this work is measured by Nokia Technologies inside either an anechoic chamber or a so-called listening room with more echo using a smartphone-shaped and -sized device that contains four microphones. The listening room in where some of the measurements were done conforms to ISO 3382-2 room echo standard and meets ITU-R BS.1116-3 NR-15 noise floor standard. An anechoic chamber is a room that is designed to absorb as much reflections as possible to minimize the echoes inside. This is done by covering the walls, floor and ceiling with non-reflective surfaces. The anechoic room as well as the rotating table can be seen in Figure 4.1 below. The anechoic room the measurements were taken in conforms to ISO 16823-1 (sound insulation for anechoic room) and ISO 3745 (free field echo properties) standards. Some datasets are pure, i.e. there's no added noise in the room, while some of them has Hoth noise or cafeteria sounds played from one or more loudspeakers. In total we test four different datasets measured in these rooms: two measured in anechoic chamber and two measured in the listening room. The data from anechoic chamber include one without noise and one with Hoth noise, and the data from listening room also include one without noise, and one with cafeteria noise played in the background.

The sampling rate of the measurement is about 47Hz, and at each time instant there are 30 values in the data, each representing a different frequency range and the values being an estimated direction of source. The directions are estimated with a proprietary algorithm for each frequency band individually. The data contains azimuth and elevation angles, time of the measurement, measured energy and a confidence value. Confidence is a value between 0 and 1, describing the confidence of the measurement. This value is used in the particle filter algorithm to lower the weights of measurements with low confidence.

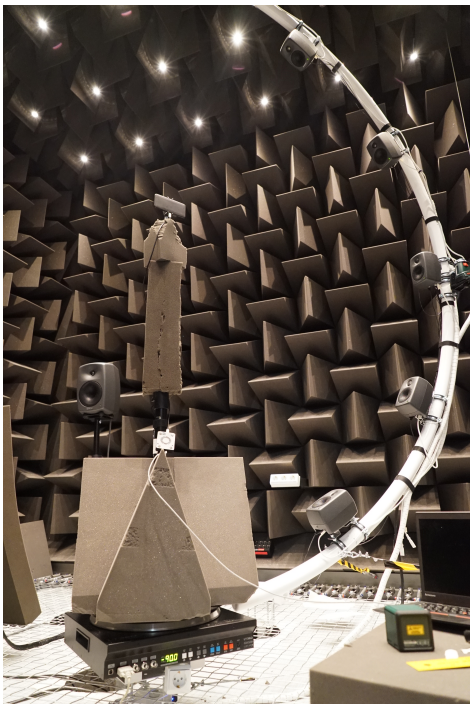

Figure 4.1. *Anechoic chamber where some of the data was recorded*

In both rooms the noise playing loudspeakers are set at $90°$ elevation. The measurement device was set on a turntable, that rotates a full circle with $1°$ accuracy at constant speed. Data recorded in the listening room are naturally more reverberant as the walls are not fully covered with non-reflective surfaces.

In some datasets the inclination of the measurement device varies between $60°$ and $120°$, which can be seen in Figure 4.2 below. The actual location in the Figure 4.2 is formed

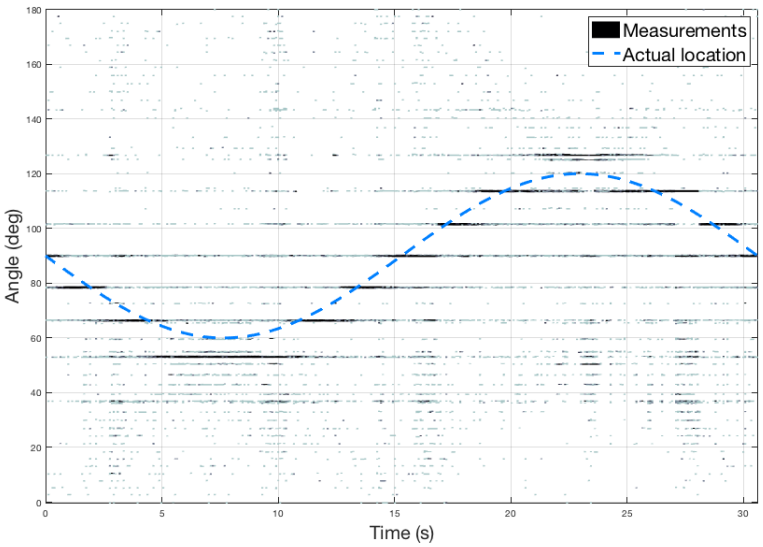

Figure 4.2. *Example of the data: elevation angle in the anechoic chamber*

from the given information of the measurements. In Figure 4.3 an example of the data can

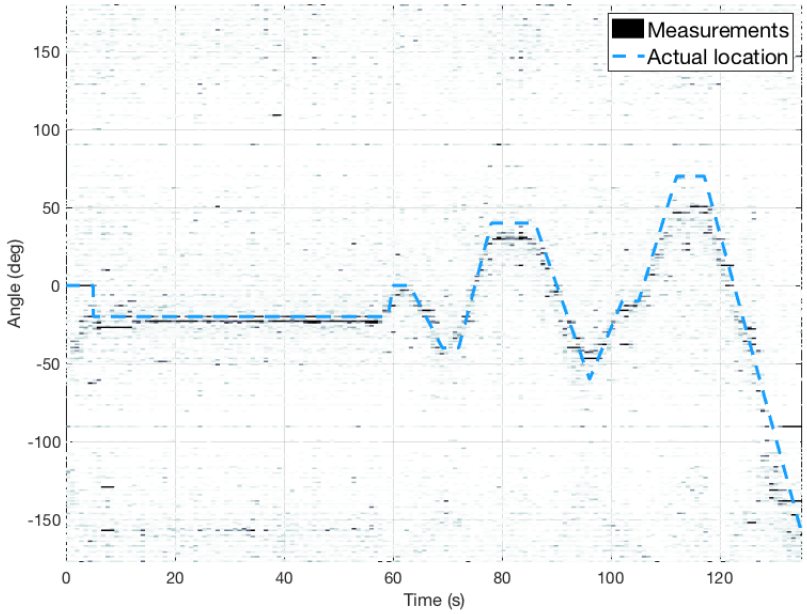

Figure 4.3. *Example of the data: azimuth angle in the listening room without noise with one speaker*

be seen, where there was one speaker in the listening room with no noise coming from loudspeakers. The speaker walked around the room randomly and the actual location is estimated from a video which accompanied the measured data. Later in Subsection 4.3.3 some randomly generated data along with real-life data, which is recorded in a racing event with cars driving by, is used.

4.2 Maximum likelihood

Maximum likelihood is used here to set a benchmark for more advanced estimations when estimating position as well as finding out the estimated value of κ in circular Gaussian or von Mises-Fisher distributions. In this section we'll see how ML was used to estimate the position and values of kappa, and what the results were.

4.2.1 Using maximum likelihood on data

ML doesn't take into account previous measurements and is done separately at each measured moment, and thus gives more inaccurate results compared to other methods. In here the ML estimation is done two different ways: first assuming that the direction is VMF distributed and then separately on each angle, assuming they have independent circular normal, or von Mises, distribution.

Results of maximum likelihood estimation of the location can be seen in Figure 4.4 and

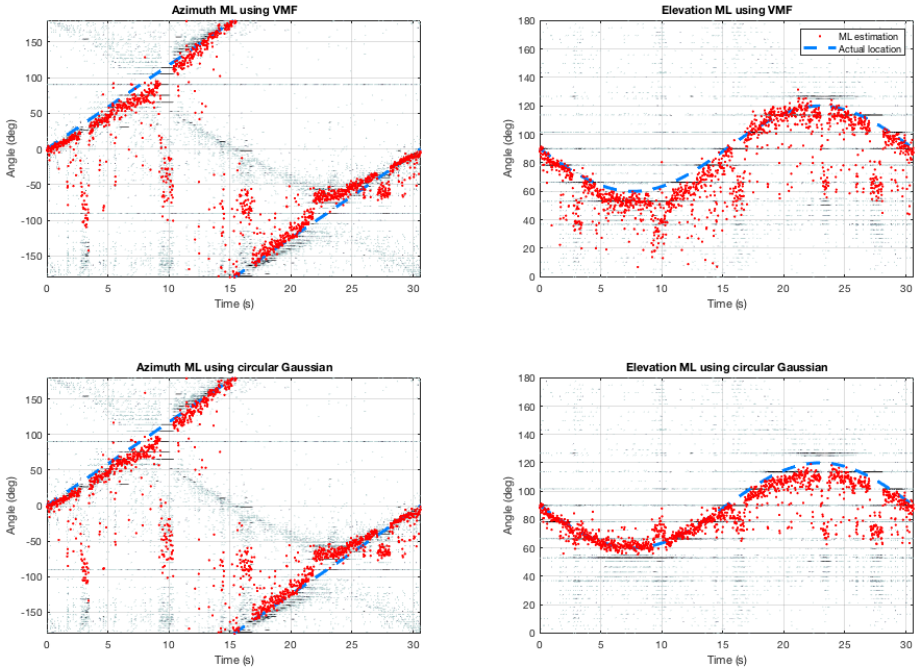

Figure 4.4. *Results of maximum likelihood estimation with different distributions*

the errors in the Table 4.1. On the angles in the top row of Figure 4.4, we have assumed the angles to be VMF distributed and the ML estimation is done on the both angles simultaneously, with the same κ value. On the bottom row we have assumed circular Gaussian distribution, and the ML estimation is done separately on the azimuth and elevation angles. As expected, the results are somewhat better when the estimation is done separately, which can be seen on the Table 4.1 below. On the table the abbreviation "CG" stands for circular Gaussian distribution and "VMF" for von Mises-Fisher distribution.

Standard deviations	Azimuth	Elevation
Anechoic, no noise, CG	$66°$	$13°$
Anechoic no noise, VMF	$65°$	$21°$
Anechoic, Hoth noise, CG	$76°$	$16°$
Anechoic, Hoth noise, VMF	$83°$	$23°$

Table 4.1. *Standard deviations of maximum likelihood estimations*

These ML results were used to find out the values of κ, as we will show next.

4.2.2 Finding out the value of κ

Maximum likelihood estimation is also used when finding out the value of κ with Equation (2.33). When using this Equation (2.33) to estimate the value of κ we need the real location of the target to calculate the average error. If the real direction is not known,

it can be determined, for example, from a provided $360°$ video where the direction of the source can be seen. This is not a totally accurate method, but gives a sufficiently good estimation. From the results shown in this chapter the real location was known for the data recorded in the anechoic chamber and the location for the data including one speaker is taken from the video.

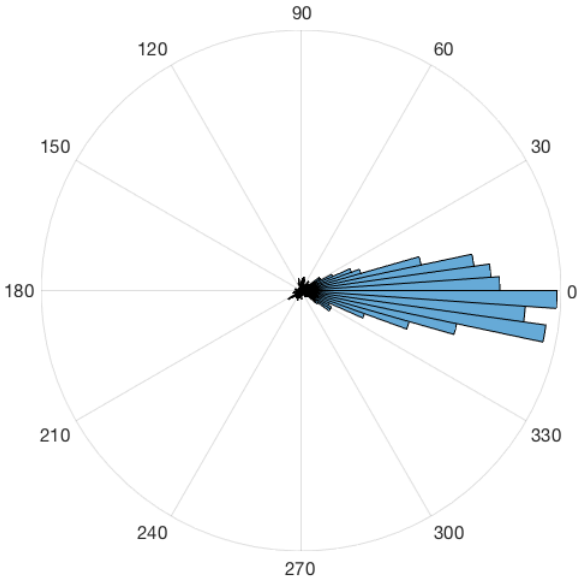

Figure 4.5. *Histogram of differences between measured and actual azimuth angles, 90 bins*

As an example, in Figure 4.5 the difference between measured and actual location of the audio source can be seen. With perfect measurements and no noise at all there would be a spike at $0°$. The angular error is calculated as a difference between the actual location and the value we got from the ML approximation (a mean value of all channels). From this calculated difference we can get the κ value by using the Equation (2.33), where now $c = 1$ and R_k are set as shown in Equation (2.23). Different values of kappas with their respective standard deviations can be found in Table 4.2. The values shown

Values of κ	κ_{azi}	σ_{azi}	κ_{ele}	σ_{ele}
Anechoic, no noise	8.7	1.9	8.6	0.8
Anechoic, Hoth noise	4.1	2.3	20.0	1.8
One speaker, no noise	6.7	2.6	20.0	1.5
One speaker, cafeteria noise	3.5	2.4	19.5	3.6

Table 4.2. *ML estimations of κ with standard deviation*

in Table 4.2 are calculated with some of the measured data ignored. This is because the measurements are assumed to have a mixed distribution between uniform and von Mises distributions. As each measurement has a given confidence value, some of the measured values can be left out of kappa analysis, as measurements with low confidence are assumed to be a part of said uniform distribution.

Estimated values of κ for elevation angles shown in the Table 4.2 seem high, especially the noisy measurements, which in a sense should be lower than the non-noisy measurements. This is because of the noise being played from the same elevation angle as the audio source is at, therefore fortifying the direction of the source. This clearly affects the results given by the particle filter on the elevation, as can be seen later.

4.3 Particle filter

When dealing with the particle filter, unlike with maximum likelihood, only circular Gaussian distribution mixed with uniform distribution is considered. This doubles the computing time as filtering has to be done separately for azimuth and elevation angles, but only 200 generated particles are used in the filtering, so the time used for calculations is satisfactory with modern devices.

As established earlier, the result relies on the used κ, but in addition to this value the results given by the particle filter depend on q-values and the used model. Next we'll go through the results given by different q-values and different models.

4.3.1 Selection of the q-value

The q-value determines the smoothness of the result we get from particle filtering. In Figure 4.6 can be seen standard deviations as a function of q-values with the constant velocity model on different datasets and how they compare against each other. A similar figure for random walk model is shown in Figure 4.7. For both models with the anechoic data the results were satisfactory until a certain threshold was reached. This threshold is more clearly seen with the random walk model in Figure 4.7. The values of κ used here are the ones presented in Table 4.2 in previous section.

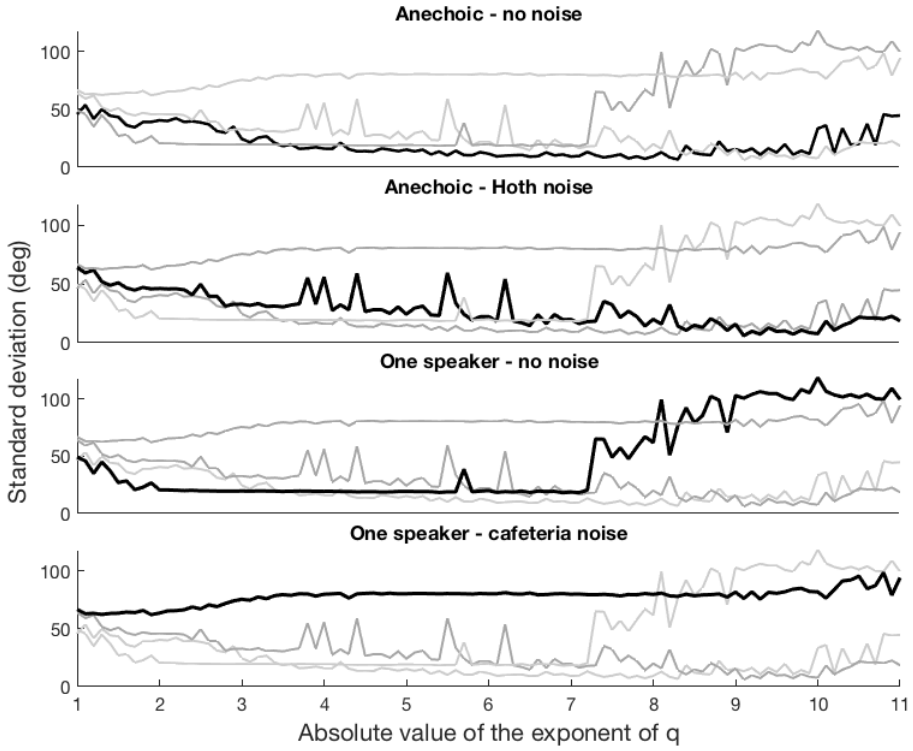

Figure 4.6. *Standard deviation vs. q-value, constant velocity model, azimuth*

Other datasets apart from the 'one speaker with cafeteria noise' have decent results, and the standard deviations are well within acceptable limits. The reason for the deficient results of one speaker with noise was later found out to be the fact that the particle filter kept on tracking the loudspeakers playing the noise in the background, instead of the speaker that was meant to be tracked. The values shown in Figure 4.6 and 4.7 can be found in Appendix A in Table A.1 and Table A.2, respectively. Some of the q-values shown in Figure 4.7 are left out from Table A.2 since after a certain value of q the results remain the same or extremely similar.

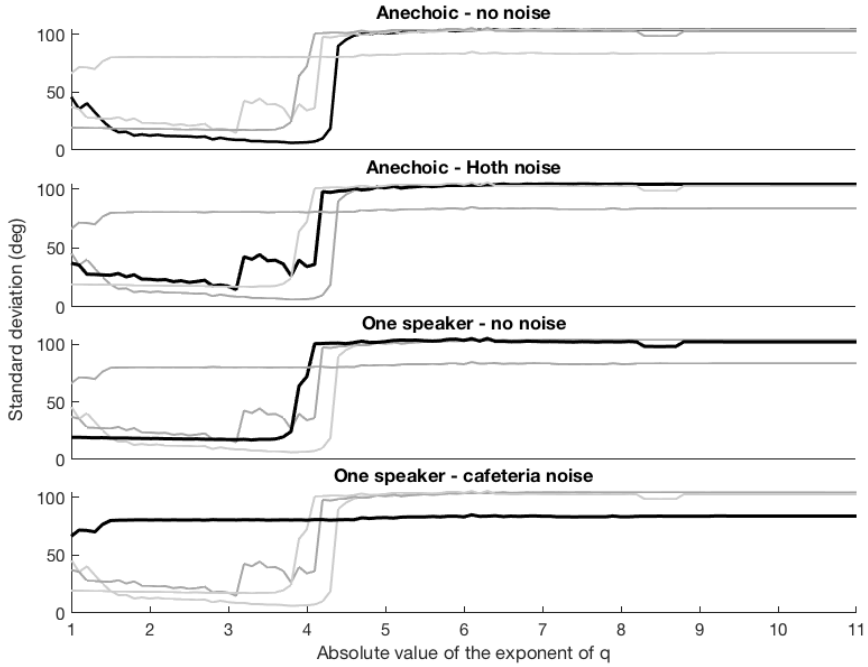

Figure 4.7. *Standard deviation vs. q-value, random walk model, azimuth*

In Figure 4.8 the results with different q-values can clearly be seen, with $q = 10^{-5.5}$ the result being too smooth and with $q = 10^{-2}$ the estimated position is not as consistent as with $q = 10^{-3}$. It is also noteworthy that with the value $q = 10^{-2}$ the standard deviation is

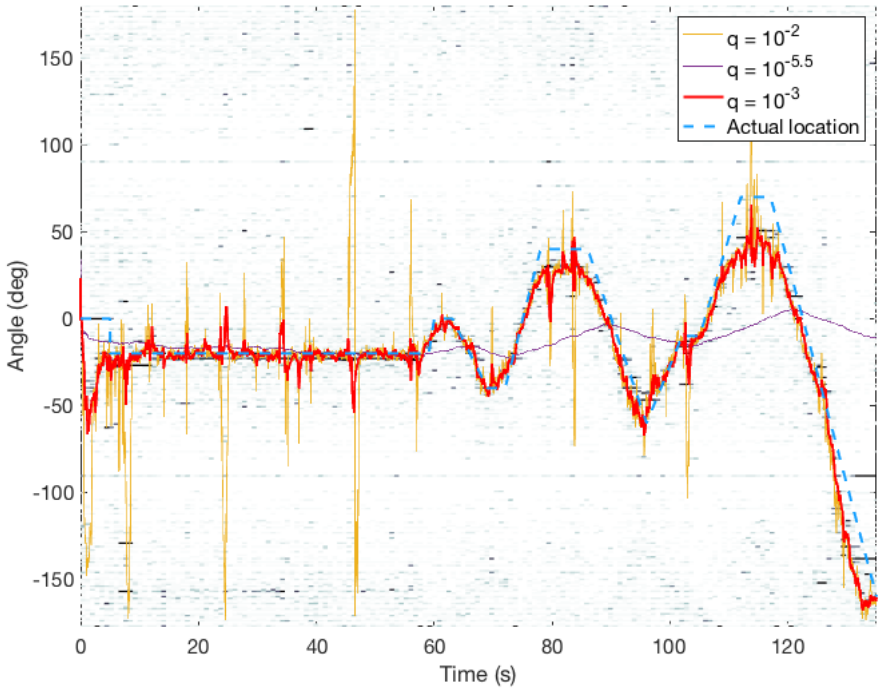

Figure 4.8. *Results of PF with different q-values, one speaker with no noise, random walk model, azimuth*

not much different from the standard deviation given by $q = 10^{-3}$ ($23.7°$ vs. $16.3°$), which is not apparent from the Figure 4.8.

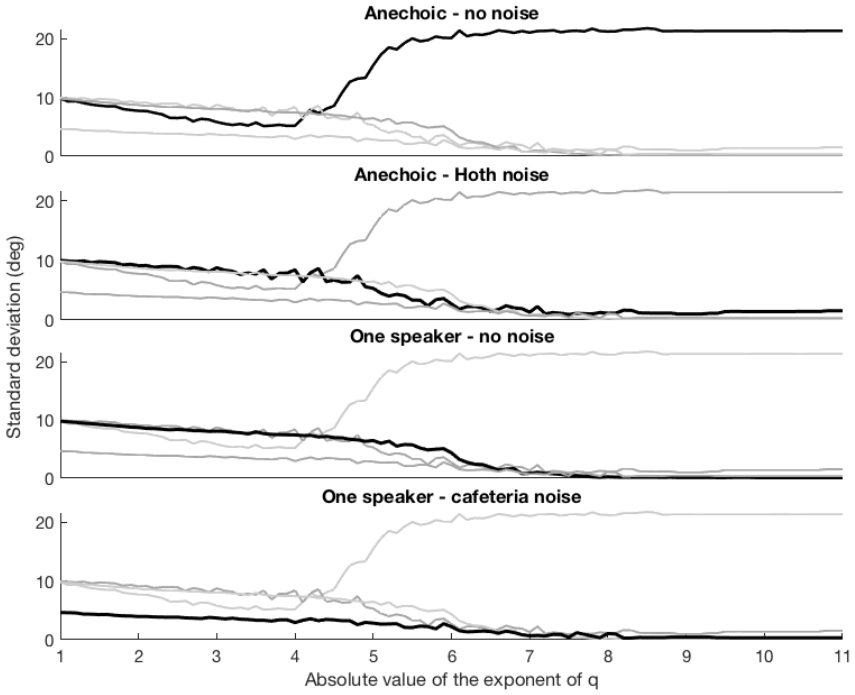

Figure 4.9. *Standard deviation vs. q-value, random walk model, elevation*

As was the case with κ values earlier, the results given by particle filter on elevation angles seem more accurate than possible given the results of the azimuth position, especially with the random walk model. These results can be seen in Figure 4.9 for the random walk model and in Figure 4.10 for the constant velocity model as standard deviation vs. used q-value. Standard deviation values for these figures can be found in Appendix A in Tables A.4 for random walk model and A.3 for constant velocity model.

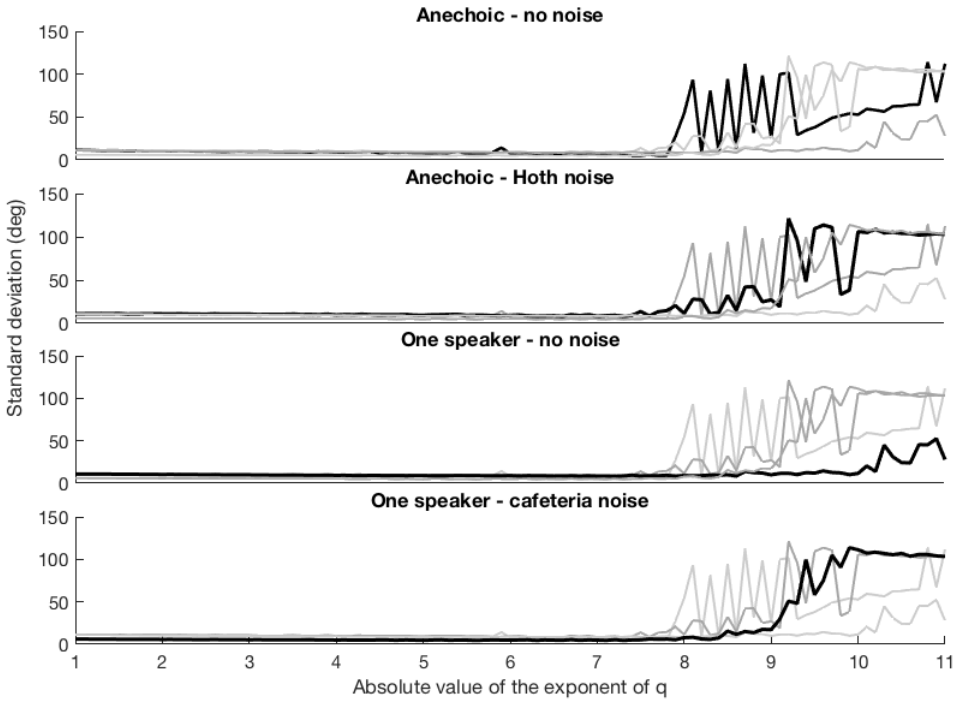

Figure 4.10. *Standard deviation vs. q-value, constant velocity model, elevation*

Reasons for the seemingly over-accurate results for elevation angle estimation are twofold. First of all, the elevation angle doesn't change a lot in the measurements and, for example with the 'one speaker' data, if we assume the elevation angle to be 90° at all times the estimation would be almost exact. Secondly, with the noisy data, the noise is played from the same elevation angle as the actual source is at, therefore only fortifying the estimation to the same elevation. Most realistic results for elevation angle estimation are with the non-noisy anechoic data, which is shown in Figure 4.11.

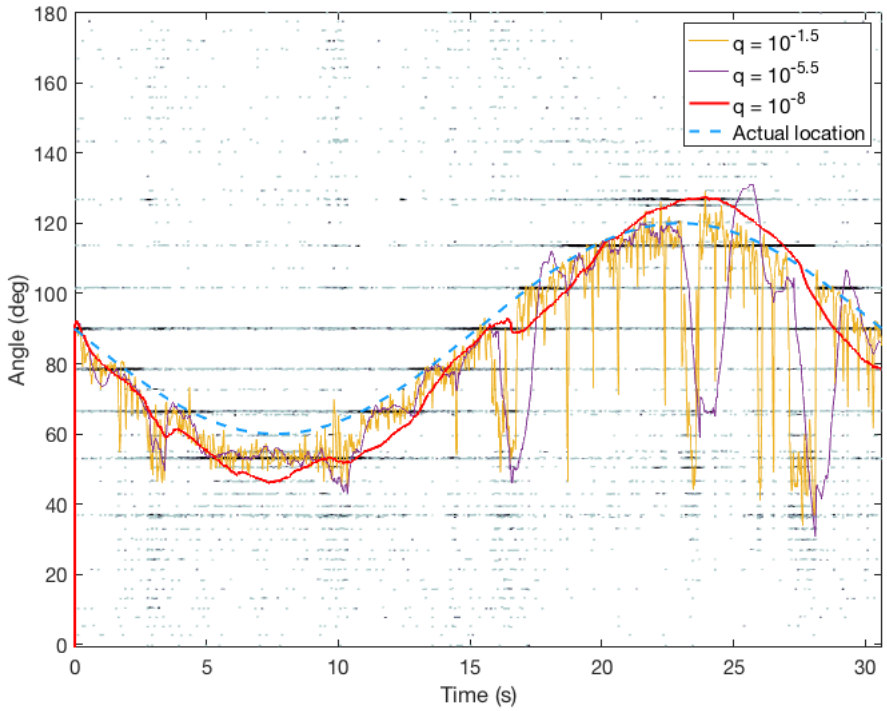

Figure 4.11. *Results of PF with different q-values, anechoic data, constant velocity model, elevation*

The dataset in question had the measurement device at an angle so that the elevation angle changes between $60°$ and $120°$ and therefore having some change in the angle.

4.3.2 Constant velocity vs. random walk models

Next up is the choice between the possible models. Earlier we proposed two possible state evolution models; random walk model represented by Equation (3.2) and a constant velocity model represented by Equation (3.7). As can be seen from the Tables A.1 and A.2, unsurprisingly, the constant velocity model works better when the audio source is moving at a constant speed (second and third columns of each table) and the random walk model works better when the source is in erratic motion (fourth and fifth columns). Most drastic difference was with the data recorded in the anechoic room including Hoth noise, where the best standard deviation of azimuth is $7.9°$ with the constant velocity model and a nearly unusable $38.3°$ with the random walk model.

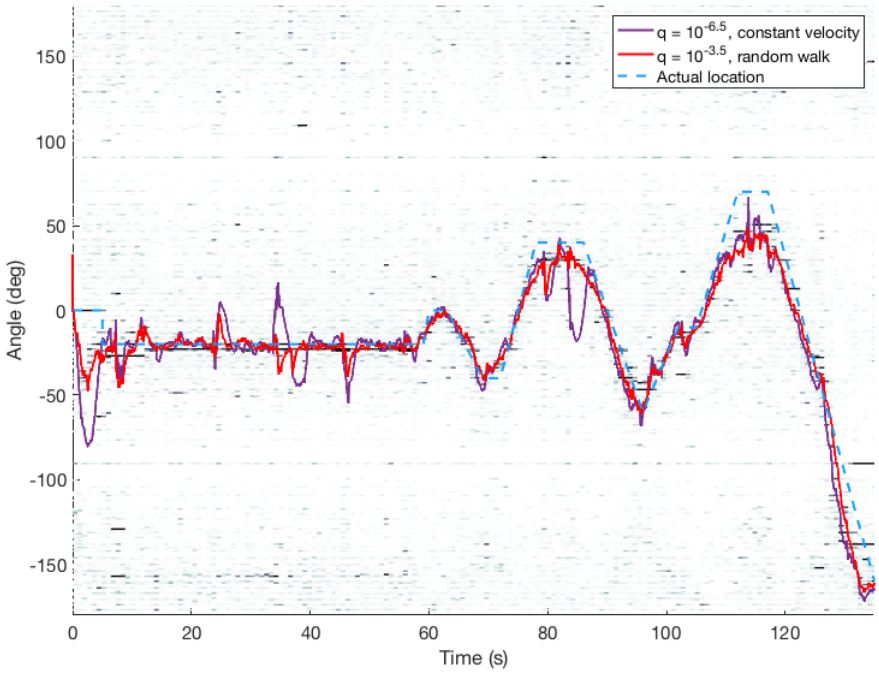

Figure 4.12. *Comparison of random walk and constant velocity models, one speaker with no noise, azimuth*

In Figure 4.12 is visually accessible information about the difference between the best results of both models on the data with one speaker who is moving freely. By assuming the free, irregular movement of the audio source the estimated angle track is smoother.

4.3.3 Tracking multiple targets with particle filter

Particle filter was also tested on tracking multiple targets, both on generated data and measured data, which was measured in a racing event where multiple cars passed the measurement device. Generated data used in this subsection was composed with constant speed model, and the same model was used in filtering as well. The distribution of particles generated by the particle filter over time can be seen in Figure 4.13, where also the generated tracks are shown. Figure 4.13 is more of an example to show the problems and possibilities in multi-target tracking than anything else. Ideally the filter would track either both of the targets or only one, the latter of which is achieved as the filtering goes on. In comparison real life data the 550 point, where the particles concentrate only on one path, would be reached after several seconds of measurements. This also causes a new problem, which is the question of how could we get another particle filter to focus on the upper path instead, which would allow us to track multiple targets with multiple filters, with the exact same data simultaneously.

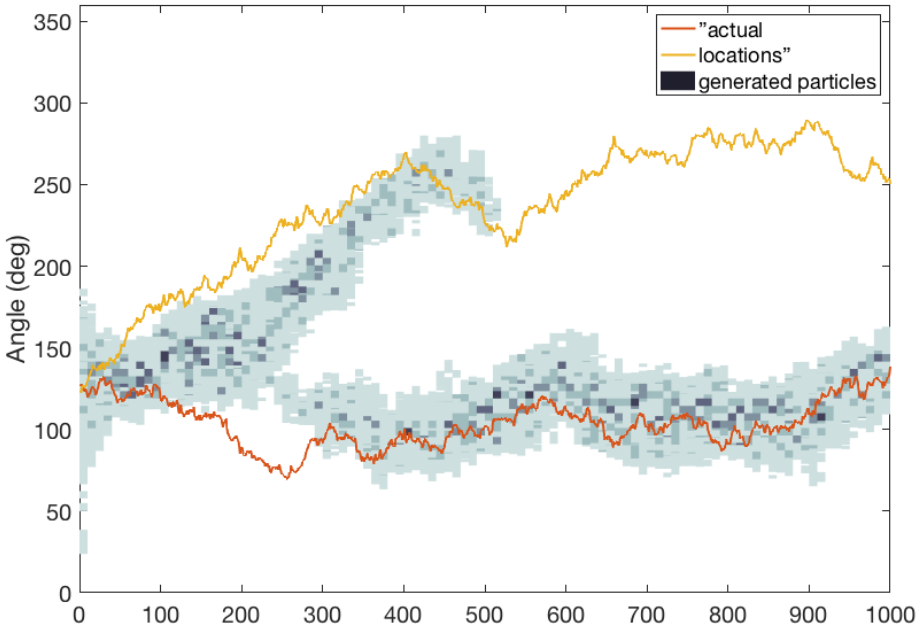

Figure 4.13. *Multiple target tracking with particle filter on generated data*

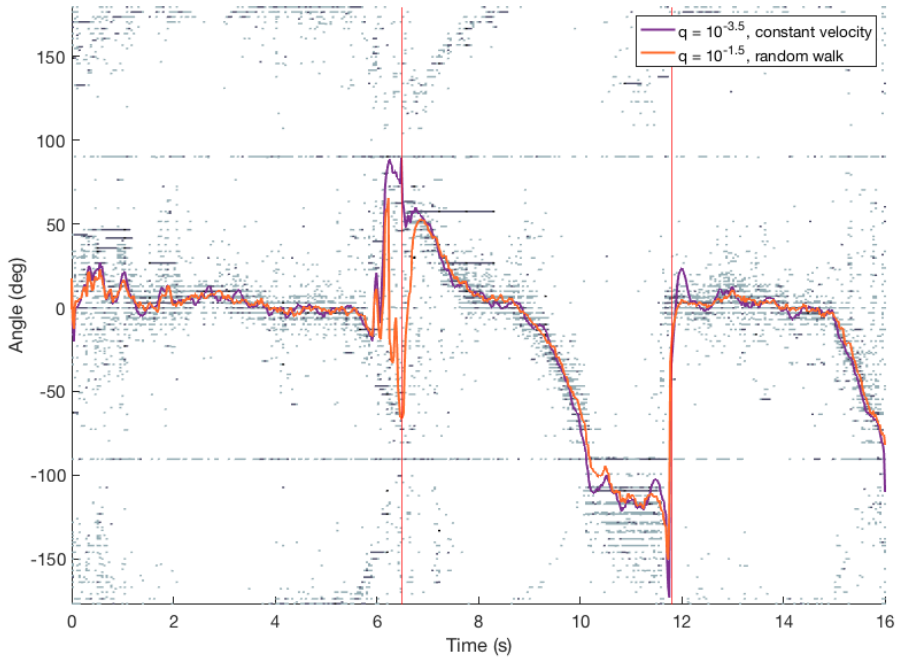

Figure 4.14. *Multiple real-life targets tracked with both models*

In Figure 4.14 is shown an example of real data with multiple targets tracked. With the red vertical lines the times where the audio source changes to another are shown. This data is recorded on a car track with vehicles driving by one by one, seen with attached position estimate in Figure 4.15. In Figure 4.15 is shown estimated azimuthal position

with vertical line while the horizontal bar represents one standard deviation of generated particles in the particle filter.

Figure 4.15. *Azimuthal position estimate fitted on the video. Vertical bar is the position estimate and horizontal bar represents one standard deviation of the generated particles from particle filter. The video was filmed by Eemi Fagerlund*

Both random walk and constant velocity models are shown in Figure 4.14, with their best respective q-values. While with either model the filter has no large problems with changing the focus to a new target, on the middle part of the data (from ~ 7 to $12s$.), although not seen in the figure, there are actually two cars and the filter focused in the middle of these which is not a wanted result. One of the weaknesses of audio-based positioning is indeed the problem of distinguishing between two different audio sources which are located in same or close to same direction. This could probably be fixed by signal processing, but that's out of scope of this thesis.

5 SUMMARY

In this work we studied particle filter and its uses in positioning of an audio source. We also developed a real-time positioning model for angle of arrival measurements, which is assuming a mixed distribution between von Mises and uniform distributions. We introduced the reader to the mathematical background of positioning and particle filter, as well as the background of estimating the κ in von Mises(-Fisher) distribution with maximum likelihood method. The particle filter in positioning was tested with real measured data, with two different motion models based on the movement of the target. We evaluated the results by comparing their respective standard deviations.

While the choice of measurement model affects the results by some amount, the biggest factor is the chosen q-value. This time the q-value was chosen by hand and tested thoroughly. This leaves the doors open for further research, as how could the value be chosen automatically to be the best without knowing the actual location of the audio source. Another large factor for the result was, of course, the amount of noise the data had. The noise had the biggest effect on the data with one speaker walking in the room, as the particle filter could not distinguish between the actual speaker and the loudspeakers playing the noise.

In addition to automating the q-value, this work leaves other open questions for future research. Not only the use of different filters, but also noting the properties of the measurement device in the model. There are certain known angles which gives worse measurements than others, which were not taken into account in this work. Also further work with the estimated position of elevation angle and noisy data is possible. The results show a positive possibility of tracking multiple targets with multiple particle filters using audio-based angle of arrival measurements; this is something that future research could focus on as well.

REFERENCES

[1] M. Abramowitz and I. Stegun. *Handbook of Mathematical Functions*. National Bureau of Standards, 1964.

[2] S. Ali-Löytty. *Kalman Filter and Its Extensions in Navigation*. Master's Thesis. Available: http://urn.fi/URN:NBN:fi:tty-201012141388. Tampere University of Technology. 2004. (Visited on 06/28/2019).

[3] Y. Ban, X. Alameda-Pineda, C. Evers and R. Horaud. Tracking Multiple Audio Sources with the von Mises Distribution and Variational EM. *IEEE Signal Processing Letters* 26(6) (2019), 798–802.

[4] A. Banerjee, I. S. Dhillon, J. Ghosh and S. Sra. Clustering on the Unit Hypersphere using von Mises-Fisher Distributions. *Journal of Machine Learning Research* 6 (2006), 1345–1382.

[5] Á. Baricz. Bounds for modified Bessel functions of the first and second kinds. *Proceedings of the Edinburgh Mathematical Society* 53 (2010), 575–599.

[6] D. J. Best and N. I. Fisher. Efficient Simulation of the von Mises Distribution. *Applied Statistics* 28(2) (1979), 152–157.

[7] V. S. Borkar, V. Ejov, J. A. Filar and G. T. Nguyen. *Hamiltonian Cycle Problem and Markov Chains*. Springer, 2012.

[8] V. Cevher, A. C. Sankaranarayanan, J. H. McClellan and R. Chellappa. Target Tracking Using a Joint Acoustic Video System. *IEEE Transactions on Multimedia* 9(4) (2007), 715–727.

[9] R. Courant and E. J. McShane. *Differential and Integral Calculus*. Blackie & Son Ltd., 1934.

[10] A. Doucet, N. de Freitas and N. Gordon. *Sequential Monte Carlo Methods in Practice*. Springer, 2013.

[11] R. Duan, K. Yang, Y. Ma, Q. Yang and H. Li. Moving source localization with a single hydrophone using multipath time delays in the deep ocean. *The Journal of the Acoustical Society of America* 136(2) (2014), EL159–EL165.

[12] N. I. Fisher, T. Lewis and B. J. J. Embleton. *Statistical Analysis of Spherical Data*. Cambridge University Press, 1993.

[13] R. Gatto and S. R. Jammalamadaka. The generalized von Mises distribution. *Statistical Methodology* 4(3) (2006), 341–353.

[14] D. B. Gennery. Visual tracking of known three-dimensional objects. *International Journal of Computer Vision* 7(3) (1992), 243–270.

[15] P. Guttorp and R. A. Lockhart. Finding the Location of a Signal: A Bayesian Analysis. *Journal of the American Statistical Association* 83(402) (1988), 322–330.

[16] T. Haubner, A. Schmidt and W. Kellermann. Active Acoustic Source Tracking Exploiting Particle Filtering and Monte Carlo Tree Search. *European Signal Processing Conference (EUSIPCO),* (2-6. Sept. 2019).

[17] A. J. Haug. *Bayesian Estimation and Tracking*. John Wiley & Sons Inc., 2012.

[18] J. D. Hol, T. B. Schön and F. Gustafsson. On Resampling Algorithms for Particle Filters. *Nonlinear Statistics Signal Processing Workshop* (2006). Available: https://ieeexplore.ieee.org/document/4378824. (Visited on 10/02/2019).

[19] L. A. Jeffress. A Place Theory of Sound Localization. *Journal of Comparative and Physiological Psychology* 41(1) (1947), 35–39.

[20] J. G. Kemeny and J. L. Snell. *Finite Markov Chains*. Springer-Verlag, 1976.

[21] M. G. Kendall. *The Advanced Theory of Statistics*. Charles Griffin & Co. LTD, 1945.

[22] M. Lefebvre. *Basic Probability Theory with Applications*. Springer, 2009.

[23] T. Li, M. Bolić and P. M. Djurić. Resampling Methods for Particle Filtering: Classification, implementation, and strategies. *Signal Processing Magazine* 32(3) (2015), 70–86.

[24] K. V. Mardia. *Statistics of Directional Data*. Academic Press, Inc., 1972.

[25] B. Ristic, S. Arulampalam and N. Gordon. *Beyond the Kalman Filter. Particle Filters for Tracking Applications*. Artech House, 2004.

[26] R. J. Rossi. *Mathematical Statistics - An introduction to Likelihood Based Inference*. John Wiley & Sons Inc., 2018.

[27] S. Särkkä. *Bayesian Filtering and Smoothing*. Cambridge University Press, 2013.

[28] J. Scott and B. Dragovic. Audio Location: Accurate Low-Cost Location Sensing. *Pervasive Computing* 3468 (2005). Available: https://doi.org/10.1007/11428572_1, 1–18.

[29] H. C. So. Efficient AoA-Based Wireless Indoor Localization for Hospital Outpatients Using Mobile Devices. *Sensors* 18(11) (2018). Available: https://doi.org/10.3390/s18113698, 3698–3715. (Visited on 05/11/2019).

[30] L. Suomalainen. *Estimating the location with angle measurements*. Master's Thesis. Available: http://urn.fi/URN:NBN:fi:tty-201705241493. Tampere University of Technology. 2017.

[31] J. Traa and P. Smaragdis. A Wrapped Kalman Filter for Azimuthal Speaker Tracking. *IEEE Signal Processing Letters* 20(12) (2013), 1257–1260.

[32] J. Traa and P. Smaragdis. Multiple speaker tracking with the Factorial von Mises-Fisher Filter. *IEEE International Workshop on Machine Learning for Signal Processing (MLSP)* (2014). doi:10.1109/mlsp.2014.6958891.

[33] J. V. Uspensky. *Introduction to Mathematical Probability*. McGraw Hill Book Company Inc., 1937.

[34] D. B. Ward, E. A. Lehmann and R. C. Williamson. Particle Filtering Algorithms for Tracking an Acoustic Source in a Reverberant Environment. *IEEE Transactions on Speech and Audio Processing* 11(6) (2003), 826–836.

A STANDARD DEVIATIONS BASED ON Q-VALUES

q-value	σ, anechoic, no noise	σ, anechoic, Hoth noise	σ, one speaker, no noise	σ, one speaker, cafeteria noise
10^{-1}	$33.6°$	$68.5°$	$57.6°$	$98.3°$
$10^{-1.5}$	$32.6°$	$60.1°$	$45.0°$	$84.8°$
10^{-2}	$39.6°$	$51.7°$	$38.4°$	$67.6°$
$10^{-2.5}$	$39.5°$	$55.2°$	$35.1°$	$67.1°$
10^{-3}	$36.6°$	$60.6°$	$33.8°$	$61.2°$
$10^{-3.5}$	$28.8°$	$51.4°$	$35.1°$	$59.5°$
10^{-4}	$39.9°$	$46.4°$	$35.5°$	$59.9°$
$10^{-4.5}$	$28.5°$	$67.6°$	$30.7°$	$59.7°$
10^{-5}	$17.8°$	$65.0°$	$27.5°$	$61.1°$
$10^{-5.5}$	$17.8°$	$30.6°$	$26.0°$	$57.7°$
10^{-6}	$13.7°$	$23.8°$	$19.2°$	$74.8°$
$10^{-6.5}$	$9.9°$	$45.1°$	$18.9°$	$71.3°$
10^{-7}	$8.5°$	$48.4°$	40.2	$62.9°$
$10^{-7.5}$	$9.6°$	$13.9°$	$81.7°$	$69.1°$
10^{-8}	$7.3°$	$22.6°$	$71.3°$	$62.4°$
$10^{-8.5}$	$13.9°$	$26.0°$	$106.9°$	$74.4°$
10^{-9}	$16.4°$	$20.5°$	$88.0°$	$72.0°$
$10^{-9.5}$	$15.1°$	$8.7°$	$95.0°$	$86.4°$
10^{-10}	$8.4°$	$7.9°$	$105.7°$	$102.5°$
$10^{-10.5}$	$15.6°$	$18.7°$	$104.8°$	$80.3°$
10^{-11}	$17.4°$	$113.6°$	$107.2°$	$90.4°$

Table A.1. *Standard deviations based on q-values, constant velocity model, azimuth.*

q-value	σ, anechoic, no noise	σ, anechoic, Hoth noise	σ, one speaker, no noise	σ, one speaker, cafeteria noise
10^{-1}	$39.2°$	$48.2°$	$32.3°$	$58.7°$
$10^{-1.5}$	$31.5°$	$47.0°$	$27.0°$	$57.8°$
10^{-2}	$23.6°$	$60.0°$	$23.7°$	$57.7°$
$10^{-2.5}$	$15.5°$	$53.2°$	$19.5°$	$53.3°$
10^{-3}	$11.4°$	$51.6°$	$16.3°$	$64.8°$
$10^{-3.5}$	$8.9°$	$45.2°$	$15.3°$	$70.9°$
10^{-4}	$10.9°$	$38.3°$	$32.5°$	$74.9°$
$10^{-4.5}$	$86.1°$	$47.0°$	$54.8°$	$76.9°$
10^{-5}	$98.6°$	$102.7°$	$92.8°$	$82.2°$
$10^{-5.5}$	$102.6°$	$104.3°$	$98.3°$	$82.8°$
10^{-6}	$103.1°$	$104.0°$	$99.6°$	$84.1°$

Table A.2. *Standard deviations of azimuth angle based on q-values, random walk model*

q-value	σ, anechoic, no noise	σ, anechoic, Hoth noise	σ, one speaker, no noise	σ, one speaker, cafeteria noise
10^{-1}	$14.6°$	$14.5°$	$15.5°$	$6.7°$
$10^{-1.5}$	$14.7°$	$14.4°$	$15.5°$	$6.6°$
10^{-2}	$14.4°$	$14.5°$	$15.4°$	$6.5°$
$10^{-2.5}$	$14.1°$	$14.5°$	$15.3°$	$6.2°$
10^{-3}	$14.2°$	$14.0°$	$15.0°$	$6.0°$
$10^{-3.5}$	$14.3°$	$14.2°$	$14.7°$	$5.8°$
10^{-4}	$15.0°$	$13.9°$	$14.7°$	$5.6°$
$10^{-4.5}$	$15.0°$	$14.1°$	$14.2°$	$5.5°$
10^{-5}	$17.1°$	$13.9°$	$14.0°$	$5.3°$
$10^{-5.5}$	$13.3°$	$13.2°$	$14.1°$	$5.1°$
10^{-6}	$11.8°$	$12.6°$	$14.0°$	$5.2°$
$10^{-6.5}$	$7.3°$	$11.8°$	$14.1°$	$5.0°$
10^{-7}	$6.8°$	$12.8°$	$15.2°$	$4.7°$
$10^{-7.5}$	$7.6°$	$30.5°$	$12.7°$	$4.5°$
10^{-8}	$5.3°$	$9.4°$	$13.9°$	$12.8°$
$10^{-8.5}$	$14.9°$	$12.1°$	$13.2°$	$4.9°$
10^{-9}	$27.8°$	$12.2°$	$22.6°$	$18.7°$
$10^{-9.5}$	$39.3°$	$103.9°$	$17.7°$	$105.3°$
10^{-10}	$57.5°$	$101.8°$	$43.4°$	$108.0°$
$10^{-10.5}$	$60.3°$	$15.8°$	$108.1°$	$7.8°$
10^{-11}	$68.4°$	$17.4°$	$108.8°$	$6.9°$

Table A.3. *Standard deviations of elevation angle based on q-values, constant velocity model*

q-value	σ, anechoic, no noise	σ, anechoic, Hoth noise	σ, one speaker, no noise	σ, one speaker, cafeteria noise
10^{-1}	$13.3°$	$13.8°$	$14.8°$	$6.0°$
$10^{-1.5}$	$12.8°$	$13.4°$	$14.4°$	$5.6°$
10^{-2}	$12.1°$	$13.1°$	$13.9°$	$5.2°$
$10^{-2.5}$	$11.2°$	$12.6°$	$13.4°$	$4.8°$
10^{-3}	$8.2°$	$12.3°$	$12.8°$	$4.5°$
$10^{-3.5}$	$5.9°$	$11.3°$	$12.3°$	$4.2°$
10^{-4}	$5.8°$	$10.3°$	$11.6°$	$3.9°$
$10^{-4.5}$	$9.9°$	$9.3°$	$10.2°$	$3.7°$
10^{-5}	$16.7°$	$7.9°$	$9.2°$	$3.5°$
$10^{-5.5}$	$18.2°$	$7.0°$	$8.1°$	$2.9°$
10^{-6}	$21.2°$	$4.4°$	$6.4°$	$2.7°$
$10^{-6.5}$	$21.1°$	$2.1°$	$4.2°$	$1.8°$
10^{-7}	$21.3°$	$1.5°$	$2.2°$	$1.0°$
$10^{-7.5}$	$21.5°$	$0.9°$	$0.6°$	$0.7°$
10^{-8}	$21.6°$	$1.0°$	$0.5°$	$0.3°$
$10^{-8.5}$	$21.4°$	$1.2°$	$0.4°$	$0.3°$
10^{-9}	$21.5°$	$1.0°$	$0.4°$	$0.3°$
$10^{-9.5}$	$21.6°$	$0.9°$	$0.3°$	$0.4°$
10^{-10}	$21.6°$	$0.9°$	$0.3°$	$0.4°$
$10^{-10.5}$	$21.6°$	$0.9°$	$0.3°$	$0.4°$
10^{-11}	$21.5°$	$0.9°$	$0.3°$	$0.4°$

Table A.4. *Standard deviations of elevation angle based on q-values, random walk model*

B MATLAB IMPLEMENTATION OF USED PARTICLE FILTER

`pfilter`: Particle filter assuming mixed distribution between von Mises and spherical uniform distribution.

Input in `pfilter`-function:

- `qk`: smoothing coefficient
- `kappa`: concentration parameter of von Mises distribution
- `meas`: measurement in degrees in $30 \times N$ matrix
- `conf`: confidence measurement in $30 \times N$ matrix
- `model`: 1 for random walk model, 2 for constant velocity

Output in `pfilter`-function:

- `pf_deg`: estimated position
- `pf_std`: standard deviations of generated particles
- `allP`: all generated particles

```
function [pf_deg, pf_std, allP] = pfilter(qk, kappa, meas, conf, model)

Delta=1;     % time step size
sigma0=0.01; % std of initial velocity
nk = length(meas);
% For replicability
rng(0);
% Wraps the input to [-pi,pi)
wrap=@(t) mod(t/pi+1,2)*pi-pi;

y = wrap(deg2rad(meas));
N=200;                % number of particles
xx=[(2*rand(1,N)-1)*pi; sigma0*randn(1,N)];

allP = nan(N,nk);
M0=mean(xx(1,:));     % estimate of initial theta
M=nan(2,nk);          % preallocate, estimates of theta
for k=1:nk
    y2 = y(:,k);
    if mean(conf(:,k)) < 0.5
        alpha = 0.4;
    else
        alpha = 0;
    end
    if model == 1
        % Random walk model
        xx(1,:) = xx(1,:)+sqrt(qk*Delta)*randn(1,N);
    elseif model == 2
        % Constant velocity model
        xx(1,:)=xx(1,:)+Delta*xx(2,:);
        xx(2,:)=xx(2,:)+sqrt(qk*Delta)*randn(1,N);
    end
    % weights from a mixed distribution
    w = (1-alpha)*exp(kappa*cos(y2(:,1)-xx(1,:)))+alpha/(2*pi);
    w = mean(w,1);
    w=w/sum(w);                          % scale the weights
    xx=xx(:,resamp(w));                  % resample
    M(1,k)=mean(xx(1,:));                % estimate of theta
    allP(:,k)=xx(1,:)';                  % all particles
    M(2,k)=std(xx(1,:));                 % the std of particles
end
allP=rad2deg(allP);
pf_deg = rad2deg(unwrap([M0 M(1,:)]));
pf_std = [std(rad2deg(M0)),rad2deg(M(2,:))];

% multinomial resampling
function J=resamp(W)
[~,J]=histc(rand(length(W),1),[0;cumsum(W(:))]);
end
end
```

Program B.1. *Used particle filter implemented in MATLAB*