Audio Source Positioning Based on Angle of Arrival Measurements Using Particle Filter

Mikko Lager, Simo Ali-Löytty, and Robert Piché
Computing Sciences
Tampere University
Tampere, Finland
simo.ali-loytty@tuni.fi

***Abstract*—This paper presents a numerical approach to the audio source positioning problem using microphones available in mobile devices. The proposed solution is a particle filter. The algorithm accuracy was tested with four different datasets, two measured in an anechoic chamber and two in a listening room. The particle filters smoothing parameter and measurement noise variance are estimated offline. The results show that this algorithm is able to position the audio source well in quiet environments.**

***Index Terms*—Audio Positioning, Particle Filter, von Mises distribution**

I. Introduction

Estimating and tracking the direction of an audio source is something that we do daily: following a speech of another person, sensing the position of a passing car 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 [1]. While human physiology makes estimating the position possible within those skills limits, computers provide a variety of other possibilities to estimate the position of an audio source, as several parameters can be measured at once and with multiple microphones. These measurements can be in the form of time of arrival, time difference of arrival, angle of arrival or received signal strength (pressure in case of audio waves), for example [2].

The ability to track an audio source has a number of useful applications. Some of the previous research on this subject include path planning for automated robots [3] and estimating positions of passing vessels in ocean environments [3], [4]. 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 [5].

In this work we focus on the use of particle filtering to track the direction of audio sources based on a sequence of instantaneous angle of arrival measurements.

The background of von Mises distribution and the estimation of its concentration parameter are discussed in Section II. In Section III we present the background and algorithm of particle filter. The used state system and measurement models are given in Section IV. Section V includes information about the used measurements. The results are discussed in Section VI and the paper is concluded in Section VII.

II. von Mises distribution and the estimation of its concentration parameter

An angle θ in the interval $(-\pi, \pi]$ is said to be von Mises distributed, denoted by $\theta \mid \phi_0, \kappa \sim \mathrm{vM}(\phi_0, \kappa)$, if its probability density function is

$$f(\theta \mid \phi_0, \kappa) = \frac{e^{\kappa \cos(\theta - \phi_0)}}{2\pi I_0(\kappa)}, \tag{1}$$

where $I_0(\kappa)$ is the modified Bessel function of first kind and order 0, ϕ_0 is the mean direction angle and $\kappa > 0$ is the concentration parameter [6, p. 48]. A conjugate prior distribution for the parameters is given by [7]

$$f(\phi_0, \kappa) \propto I_0(\kappa)^{-c} e^{\kappa R_0 \cos(\phi_0 - \lambda)}, \tag{2}$$

with hyperparameters $R_0 \geq 0$, $c \geq 0$, $\lambda \in (-\pi, \pi]$. Given angle measurements $\theta_1, \ldots, \theta_k$, conditionally independent given ϕ_0, κ, the posterior distribution is

$$f(\phi_0, \kappa | \theta_{1:k}) \propto \frac{1}{I_0(\kappa)^{c+k}} e^{\kappa R_k \cos(\phi_0 - \phi_k)}, \tag{3}$$

where ϕ_k and R_k are obtained from

$$\begin{aligned} R_k \cos(\phi_k) &= R_0 \cos(\lambda) + \sum_{i=1}^{k} \cos(\theta_i), \text{ and} \\ R_k \sin(\phi_k) &= R_0 \sin(\lambda) + \sum_{i=1}^{k} \sin(\theta_i). \end{aligned} \tag{4}$$

For offline estimation with known angle of arrival data we can assume $\phi_0 = 0$, and the maximum a posteriori estimate of the concentration parameter is

$$\kappa = A^{-1}\left(\frac{R_k}{c+k}\right), \tag{5}$$

where $A(\kappa) = \frac{I_1(\kappa)}{I_0(\kappa)}$. We compute A^{-1} using a numerical code for finding a root in an interval. The interval's limits are set using Theorem 1 of [8].

This research was funded by Nokia Technologies.

978-1-7281-6455-7/20/$ 31.00 © 2020 IEEE

III. Particle filter

The particle filter is a Monte Carlo method that approximates the posterior distribution with particles, weighted samples of the state vector.

In a Monte Carlo simulation we have N independent random samples of the posterior distribution of the state x_k at time k given measurements $\{y_1, \ldots, y_k\}$:

$$x_k^{(i)} \sim f\left(x_k \mid y_{1:k}\right),\ i \in \{1, \ldots, N\}. \tag{6}$$

Then the expectation

$$\mathrm{E}\left[g\left(x_k\right) \mid y_{1:k}\right] = \int g\left(x_k\right) f\left(x_k \mid y_{1:k}\right) dx_k, \tag{7}$$

where g is an arbitrary integrable function, can be approximated as [9, 116-117]

$$\mathrm{E}\left[g\left(x_k\right) \mid y_{1:k}\right] \approx \frac{1}{N} \sum_{i=1}^{N} g(x_k^{(i)}). \tag{8}$$

In a filtering context, we can't directly sample the posterior, but from Bayes' law we know that its density is proportional to the product of the measurement density $f\left(y_k \mid x_k\right)$ and the prior density $f\left(x_k \mid x_{k-1}\right)$. So, we can approximate the expectation using importance sampling. Instead of (6), we sample from an importance distribution

$$x_k^{(i)} \sim \pi\left(x_k \mid y_{1:k}\right),\ i \in \{1, \ldots, N\}. \tag{9}$$

Now, (7) can be rewritten as

$$\begin{aligned} \mathrm{E}\left[g\left(x_k\right) \mid y_{1:k}\right] &= \int g\left(x_k\right) f\left(x_k \mid y_{1:k}\right) dx_k \\ &= \int \left[g\left(x_k\right) \frac{f\left(x_k \mid y_{1:k}\right)}{\pi\left(x_k \mid y_{1:k}\right)}\right] \pi\left(x_k \mid y_{1:k}\right) dx_k, \end{aligned} \tag{10}$$

This is the expectation of the bracketed term over the importance distribution, which allows us to form a Monte Carlo approximation. After generating the samples we calculate the weights

$$w^{*(i)} = \frac{f\left(y_k \mid x_k^{(i)}\right)}{\pi\left(x_k^{(i)} \mid y_{1:k}\right)} f(x_k^{(i)} \mid y_{1:k-1}), \tag{11}$$

and normalize them,

$$w^{(i)} = \frac{w^{*(i)}}{\sum_{j=1}^{N} w^{*(j)}}. \tag{12}$$

The posterior expectation can now be formed with these normalized weights as [9, 118-120]

$$\mathrm{E}\left[g\left(x\right) \mid y_{1:k}\right] \approx \sum_{i=1}^{N} w^{(i)} g\left(x^{(i)}\right) \tag{13}$$

Sequential Monte Carlo methods exhibit *degeneracy*, whereby almost all the weight eventually becomes concentrated in a few particles. This can be solved by periodically resampling [10], [11]: low weight particles are eliminated and high weight particles are replicated; then the weights are reset to $1/N$. In multinomial resampling, the new particles are $x^{(m_1)}, \ldots, x^{(m_N)}$, where the indices $m_i \in \{1, \ldots, N\}$ are independent samples from the categorical distribution having probability mass function $w_{1:N}$.

The basic particle filter, the so called bootstrap filter, uses the filter state distribution $f(x_k|x_{k-1}^{(i)})$ as the importance distribution, so that (11) reduces to

$$w^{*(i)} = f(y_k \mid x_k^{(i)}), \tag{14}$$

and resampling is done at every time step. The algorithm is outlined below.

Algorithm 1 Particle Filter

1: Generate N samples $x_0^{(i)}$ from the initial state distribution $f(x_0)$
2: Set weights $w_0^{(i)} = \frac{1}{N}$ for $i = 1, \ldots, N$
3: **for** $k = 1 : T$ **do**
4: Generate samples $x_k^{(i)}$ from the state model distribution $f(x_k|x_{k-1}^{(i)})$
5: For each channel j, calculate particle weights $w_{k,j}^{*(i)}$ for frequency band measurement $y_{k,j}$ using (14)
6: Set $w_k^{*(i)}$ = the confidence-weighted mean of weights $w_{k,:}^{*(i)}$
7: Normalize particle weights using (12)
8: Resample
9: Calculate the filter state mean $\bar{x}_k$ using (13) with $g =$ the identity function.
10: **end for**

Note that step 6 is a heuristic data-binning step that is not part of the standard particle filter. Also, for state components that represent angles, the circular mean is used in place of the arithmetic mean.

IV. Models

A. State system models

In this work we tested two different state evolution models: random walk and constant velocity models. In the random walk model the state is the azimuth angle

$$x = y_{\text{azi}} \in (-\pi, \pi] \tag{15}$$

and the state evolution model is

$$x_k = x_{k-1} + q_{k-1}, \tag{16}$$

where $q_{k-1} \sim N(0, q\Delta_t)$ is a normally distributed error with mean 0 and variance $q\Delta t$; $\Delta_t = t_k - t_{k-1}$ is the constant time step length.

In the constant velocity model the state is the tuple

$$x = \begin{bmatrix} y_{\text{azi}} \\ v \end{bmatrix}, \tag{17}$$

where v is the angular velocity. The state evolution model is

$$x_k = F x_{k-1} + q_{k-1}. \tag{18}$$

with state transition matrix

$$F = \begin{bmatrix} 1 & \Delta t \\ 0 & 1 \end{bmatrix}, \tag{19}$$

and driving noise $q_{k-1} \sim N(0, Q_{k-1})$ with covariance matrix

$$Q_{k-1} = q \begin{bmatrix} \frac{\Delta_t^3}{3} & \frac{\Delta_t^2}{2} \\ \frac{\Delta_t^2}{2} & \Delta_t, \end{bmatrix}. \tag{20}$$

In both state models the smoothness of the motion is governed by the q parameter. [12]

B. Measurement model

In this work the measurements are noisy azimuth angles, modelled as a mixture of von Mises and uniform distributions

$$f(y_k \mid x_{1,k}) = (1-\alpha)\frac{e^{\kappa \cos(y_k - x_{1,k})}}{2\pi I_0(\kappa)} + \frac{\alpha}{2\pi}. \tag{21}$$

The mixture parameter $\alpha \in [0, 1)$ governs the prevalence of outliers.

V. Measurements

The real data used in this work was measured by Nokia Technologies inside either an anechoic chamber (ISO 16283-1, ISO 3745) or a so-called listening room (ITU-R BS.1116-3 NR-15) with more echo using a smartphone-shaped and -sized device that contains four microphones. Some datasets are pure, i.e. there's no added noise in the room, while two of them have Hoth noise or noise recorded in a public space played from one or more loudspeakers:

1) anechoic chamber with Hoth noise,
2) anechoic chamber with no noise,
3) listening room with no noise or
4) listening room with additional noise.

In the dataset recorded in anechoic chamber without noise the measurement device was set to an angle so that the elevation angle changes between $60°$ and $120°$, while in other measurements the elevation angle is relatively constant.

The sampling rate of the measurement is about $47\,\text{Hz}$, and at each time instant there are 26 values in the data, each representing a different frequency range and the values being an estimated source direction. The directions were estimated for each frequency band individually using a proprietary algorithm.

The data is a time series of azimuth and elevation angles, time of the measurement, and a confidence value between 0 and 1. This value was used in the particle filter algorithm to lower the weights of measurements with low confidence. The weight for each particle is calculated for all twenty-six channels based on the measurement's confidence, from which a mean weight is set for the particle weight. Confidence is also used to calculate the parameter α in (21) so that it is corresponding to the percentage of low confidence measurements.

In both rooms the noise-playing loudspeakers were set at $90°$ elevation, i.e. the same horizontal plane as the measurement device. The measurement device was set on a turntable that rotates a full circle with $1°$ accuracy at constant speed.

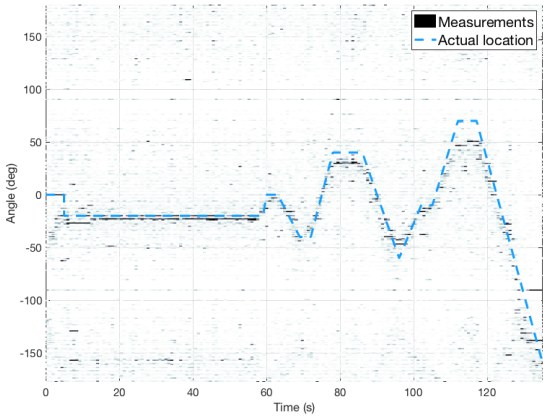

Fig. 1. Example of the data: azimuth angle in the listening room without noise with one speaker

Data recorded in the listening room are naturally more reverberant as the walls are not fully covered with non-reflective surfaces.

In addition to measured data, the listening room measurements were accompanied by a $360°$ video for annotation purposes. From this video some of the actual locations of the talker in the listening room were estimated as the actual location was not otherwise known. The actual location of the audio source in the anechoic chamber was known.

In Fig. 1 an example of the data can be seen, where there was one human speaker in the listening room with no additional 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.

VI. Results

A. Maximum likelihood on κ estimation

We used maximum likelihood estimation (ML) to determine the value of κ with (5). When using (5) to estimate the value of κ we need the real location of the target to calculate the average error. If the real direction was not known, it was determined from a provided $360°$ video where the direction of the source can be seen. This is not a totally accurate method, but it gave a sufficiently good estimation.

The angular error was 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 (5), where now $c = 1$, and R_k are set as shown in (4) with $R_0 = 0$. Different values of kappas with their respective standard deviations can be found in Table I.

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

TABLE I
MAXIMUM LIKELIHOOD ESTIMATIONS OF κ AND THEIR RESPECTIVE STANDARD DEVIATIONS

Values of κ	κ	σ_{azi}
Anechoic, no noise	8.7	1.9
Anechoic, Hoth noise	4.1	2.3
One speaker, no noise	6.7	2.6
One speaker, additional noise	3.5	2.4

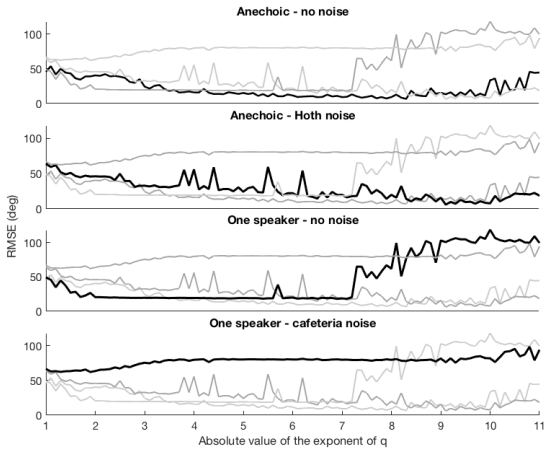

Fig. 2. RMSE vs. q-value, constant velocity model. In every sub-figures contains the same four datasets with the relevant one in black.

measurements with low confidence are assumed to be a part of said uniform distribution.

B. Particle filter

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. The q-value determines the smoothness of the result we get from particle filtering. Next we'll go through the results given by different q-values and different models. In Particle filter we use 200 particles, we found that this number of particles is enough for this application.

1) Selection of the smoothing parameter: In Fig. 2 can be seen RMS-errors 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 the random walk model is shown in Fig. 3. 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 Fig. 3. The κ values used here are the ones presented in Table I in previous section.

Other datasets apart from the 'one speaker with additional noise' have decent results, and the RMSEs are well within acceptable limits. The reason for the deficient results of one speaker with noise was later found to be the fact that the particle filter tracked the loudspeakers playing the noise in the background, instead of the human speaker that was meant to be tracked. In this case, the signal-to-noise ratio (SNR) was low and the noise field was difficult for audio source tracking. The ambient sound field was simulated using multiple loudspeakers, which may not fully correspond to a realistic noise field. Furthermore, the recorded noise also contains speech. Consequently, part of the direction estimates reflect the background noise instead of the desired sound source, which makes sound source tracking difficult.

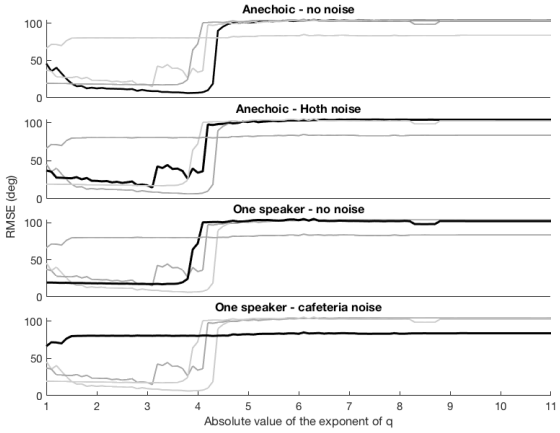

Fig. 3. RMSE vs. q-value, random walk model. In every sub-figures contains the same four datasets with the relevant on in black.

In Fig. 4 the effects of 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 $q = 10^{-2}$

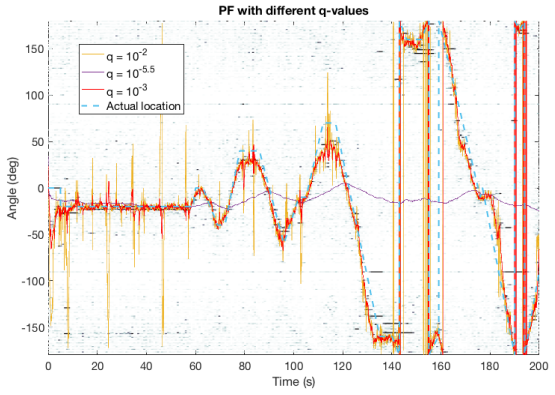

Fig. 4. Results of PF with different q-values, one speaker with no noise, random walk model

the RMSE is not much different from the RMSE given by $q = 10^{-3}$ (23.7° vs. 16.3°), which is not apparent from the Fig. 4.

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 (16) and a constant velocity model represented by (18). Unsurprisingly, the constant velocity model worked better when the audio source is moving at a constant speed and the random walk model works better when the source is

in erratic motion. The most drastic difference was with the data recorded in the anechoic room including Hoth noise, where the best standard deviation is 7.9° with the constant velocity model and a nearly unusable 38.3° with the random walk model.

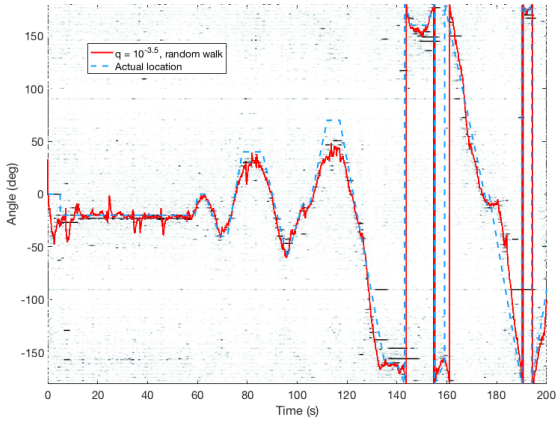

Fig. 5. Random walk model, one speaker with no noise

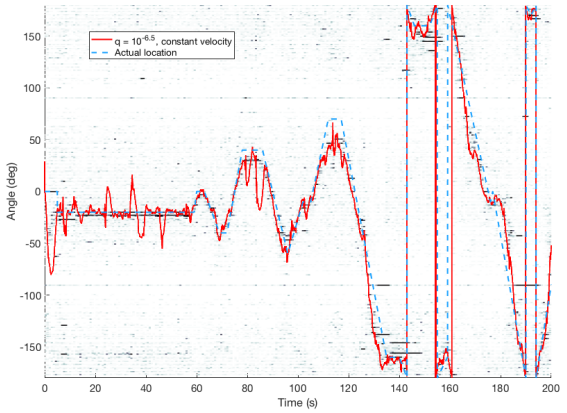

Fig. 6. Constant velocity model, one speaker with no noise

In Fig.5 and 6 is shown 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. Results of both models on different q-values can be found in the Appendix I in Tables II and III.

VII. Conclusion

In this paper we have presented the study and implementation of a particle filter to track the angular position of an audio source. The particle filter in positioning was tested with real measured data, with two different motion models based on the movement of the target.

While the choice of measurement model affects the results by some amount, the biggest factor is the chosen smoothness parameter q. Here the q-value was chosen by hand and tested thoroughly. Another large factor for the result was, of course, the amount of noise in the data. 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 additional noise. This dataset represents difficult conditions in which the SNR is low and the background noise also contains speech.

References

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

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

[3] 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.

[4] 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*, vol. 136(2), pp. EL159–EL165, 2014.

[5] 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*, vol. 26(6), pp. 798–802, 2019.

[6] N. I. Fisher, *Statistical analysis of circular data*. Cambridge University Press, 1993.

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

[8] T.-K. Chang, S. Chen, and A. Mehta, "Localization algorithm with circular representation in 2d and its similarity to mammalian brains," 2018. Available: https://arxiv.org/abs/1809.02910.

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

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

[11] 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.

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

Acknowledgements

This research was funded by Nokia Technologies. Our thanks to Nokia Technologies for providing the datasets used in this study.

APPENDIX A
RMS ERRORS BASED ON Q-VALUES

TABLE II
RMSE BASED ON Q-VALUES, CONSTANT VELOCITY MODEL, AZIMUTH.

q-value	RMSE, anechoic, no noise	RMSE, anechoic, Hoth noise	RMSE, one speaker, no noise	RMSE, 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 III
RMSE OF AZIMUTH ANGLE BASED ON Q-VALUES, RANDOM WALK MODEL

q-value	RMSE, anechoic, no noise	RMSE, anechoic, Hoth noise	RMSE, one speaker, no noise	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°