Applying von Mises Distribution to Microphone Array Probabilistic Sensor Modelling.

Ivan Marković, Ivan Petrović
University of Zagreb Faculty of Electrical Engineering and Computing, Croatia

Abstract

This paper deals with the problem of localizing and tracking a moving speaker over the full range around the mobile robot. The proposed algorithm is based on estimating the time difference of arrival by maximizing the weighted cross-correlation function in order to, by series of geometrical calculation, determine the azimuth angle of the detected speaker. A post processing technique is proposed in which each of these microphone-pair determined azimuths are further combined into a mixture of the von Mises distributions, thus producing a practical probabilistic representation of the microphone array measurement. It is shown that this distribution is inherently multimodal and that the system at hand is non-linear, which required a discrete representation of the distribution function by means of particle filtering. Experiments were conducted and the results show that the algorithm can reliably and accurately localize and track a moving sound source.

1 Introduction

In biological lifeforms hearing, as one of the traditional five senses, elegantly supplement other senses as being omnidirectional, not limited by physical obstacles, and absence of light. Inspired by these unique properties, researchers strive towards endowing mobile robots with auditory systems to further enhance human–robot interaction, not only by means of communication but also, just as humans do, to make intelligent analysis of the surrounding environment. By providing speaker location to other mobile robot systems, like path planning, speech and speaker recognition, such system would be a step forward in developing a fully functional human–aware mobile robots.

An auditory system must provide accurate estimates and must be updated frequently in order to be useful in practical tracking applications. Furthermore, the estimator must be computationally non-demanding and possess a short processing latency to make it practical for real-time systems.

Existing speaker localization strategies can be categorized in four general groups: those based upon maximizing the steered response power of a beamformer [1, 2], techniques adopting high-resolution spectral estimation concepts [3], physiologically inspired approaches [4, 5] and approaches employing Time Difference of Arrival (TDOA) information [6, 7].

Even though the TDOA estimation methods are outperformed to a certain degree by several more elaborate methods [8, 9], they still prove to be extremely effective due to their elegance and low computational costs. This paper proposes a new speaker localization method based on TDOA estimation using a microphone array of 4 omnidirectional microphones. The proposed algorithm is based on probabilistic modelling of the microphone pair measurements and particle filtering, which enables us to solve the front-back ambiguity, increase the robustness by using all the available measurements, and to localize and track speaker over the full range around the mobile robot. The main contribution of this paper is the proposed sensor model to be used for *a posteriori* inference about the location of the sound source.

One should note, however, that the proposed sensor modelling with von Mises distribution is independent of the azimuth estimation method and, furthermore, can be just as easily applied to other sensor systems, like (omnidirectional) vision, laser scans etc.

The rest of the paper is organized as follows. Section 2 presents the TDOA estimation algorithm and how this information is used to calculate the speaker azimuth. Section 3 defines the general framework for the particle filtering algorithm and introduces the von Mises distribution and the proposed measurement model. Section 4 presents experiments. In the end, Section 5 concludes the paper and presents future work.

2 TDOA Estimation

The main idea behind TDOA-based locators is a two step one. Firstly, TDOA estimation of the speech signals relative to pairs of spatially separated microphones is performed. Secondly, this data is used to infer about speaker location. The TDOA estimation algorithm for 2 microphones is described first.

2.1 Principle of TDOA

In order to determine the delay $\Delta\tau_{ij}$ in the signal captured by two different microphones (i and j), it is necessary to define a coherence measure which will yield an explicit global peak at the correct delay. Cross-correlation is the most common choice, since we have at two spatially separated microphones (in an ideal homogeneous, dispersion-free and lossless scenario) two identical time-shifted signals. Cross-correlation is defined by the following expression:

$$R_{ij}(\Delta\tau) = \sum_{n=0}^{L-1} x_i[n]\, x_j[n-\Delta\tau], \tag{1}$$

where x_i and x_j are the signals received by microphone i and j, respectively. As stated earlier, R_{ij} is maximal when $\Delta\tau$, correlation lag in samples, is equal to the delay between the two received signals.

In order to significantly lower the computational intensity of the algorithm, cross-correlation is estimated in the frequency domain:

$$\hat{R}_{ij}(\Delta\tau) = \sum_{k=0}^{L-1} \psi(k) X_i(k) X_j^*(k) e^{j2\pi\frac{k\Delta\tau}{L}}, \tag{2}$$

where $X_i(k)$ and $X_j(k)$ are the discrete Fourier Transforms (DFTs) of $x_i[n]$ and $x_j[n]$, $\psi(k)$ is a generalized weighting function, and $(.)^*$ denotes complex-conjugate. We are windowing the frames with rectangular window and no overlap. Therefore, before applying Fourier transform to signals x_i and x_j, it is necessary to zero-pad them with at least L zeros, since we want to calculate linear, and not circular convolution.

A major limitation of the cross-correlation given by (2) is that the correlation between adjacent samples is high, which has an effect of wide cross-correlation peaks. Therefore, appropriate weighting should be used.

2.2 Spectral weighting

The problem of wide peaks in unweighted, i.e. generalized, cross-correlation (GCC) can be solved by whitening the spectrum of signals prior to computing the cross-correlation. The most common weighting function is the Phase Transform (PHAT) which, as it has been shown in [10], under certain assumptions yields MLE. What PHAT function ($\psi_{\text{PHAT}} = 1/|X_i(k)||X_j^*(k)|$) does, is that it whitens the cross-spectrum of signals x_i and x_j, thus giving a sharpened peak at the true delay.

Using just the PHAT weighting poor results were obtained and we concluded that the effect of the PHAT function should be tuned down. As it was explained and shown in [11], the main reason for this approach is that speech can exhibit both wide-band and narrow-band characteristics. For example, if uttering the word "shoe", "sh" component acts as a wide-band signal and voiced component "oe" as a narrow-band signal.

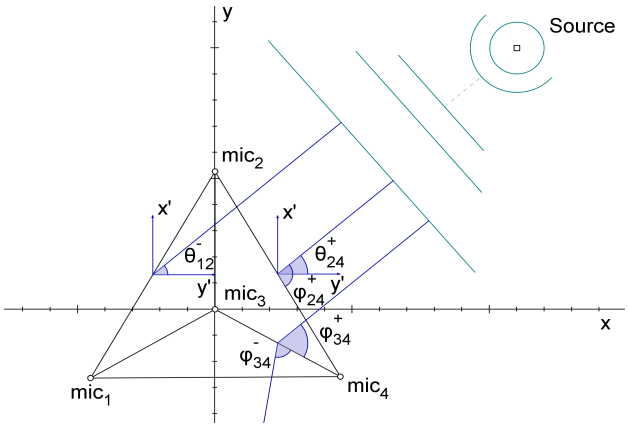


Figure 1: DOA angle transformation

Based on the discussion above, the GCC-PHAT-β has the following form:

$$\hat{R}_{ij}^{\text{PHAT-}\beta}(\Delta\tau) = \sum_{k=0}^{L-1} \frac{X_i(k)X_j^*(k)}{(|X_i(k)||X_j(k)|)^{\beta}} e^{j2\pi\frac{k\Delta\tau}{L}}. \tag{3}$$

where $0 < \beta < 1$ is the tuning parameter.

2.3 Direction of Arrival estimation

The TDOA between microphones i and j, $\Delta\tau_{ij}$, can be found by locating the peak in the cross-correlation:

$$\Delta\tau_{ij} = \arg\max_{\Delta\tau} \hat{R}_{ij}^{\text{PHAT}-\beta}(\Delta\tau). \tag{4}$$

Once TDOA estimation is performed, it is possible to compute the position of the sound source through series of geometrical calculations. It is assumed that the distance to the source is much larger than the array aperture, i.e. we assume the so called far-field scenario. Thus, the expanding acoustical wavefront is modelled as a planar wavefront. Using the cosine law we can state the following:

$$\varphi_{ij} = \pm \arccos\left(\frac{c\Delta\tau_{ij}}{a_{ij}}\right), \tag{5}$$

where a_{ij} is the distance between the microphones, c is the speed of sound, and φ_{ij} is the Direction of Arrival (DOA) angle.

Since we will be using more than two microphones one must make the following transformation in order to fuse the estimated DOAs. Instead of measuring the angle φ_{ij} from the baseline of the microphones, transformation to azimuth θ_{ij} measured from the x axis of the array coordinate system (bearing line is parallel with the x axis when

$\theta_{ij} = 0°$) is performed. The transformation is done with the following equation (angles φ_{24}^{+} and θ_{24}^{+} in **Figure 1**):

$$\begin{aligned}\theta_{ij}^{\pm} &= \alpha_{ij} \pm \varphi_{ij} \\ &= \operatorname{atan2}\left(\frac{y_j - y_i}{x_j - x_i}\right) \pm \arccos\left(\frac{c\Delta\tau_{ij}}{a_{ij}}\right). \quad (6)\end{aligned}$$

The framework for TDOA and DOA estimation was presented in detail in [6].
At this point one should note the following:

- under the far-field assumption, all the DOA angles measured anywhere on the baseline of the microphones are equal, since the bearing line is perpendicular to the expanding planar wavefront (angles θ_{12}^{-} and θ_{24}^{+} in Figure 1)
- front-back ambiguity is inherent when using only two microphones (angles φ_{34}^{-} and φ_{34}^{+} in Figure 1).

Having M microphones, (6) will yield $2 \cdot \binom{M}{2}$ possible azimuth values. How to solve the front-back ambiguity and fuse the measurements is explained in Section 3.

3 Speaker Localization and Tracking

The problem at hand is to analyse and make inference about a dynamic system. For that, two models are required: one describing the evolution of the speaker's state over time (system model), and second relating the noisy measurements to the speaker's state (measurement model). We assume that both models are available in probabilistic form. Thus, the approach to dynamic state estimation consists of constructing the *a posteriori* pdf of the state based on all available information, including the set of received measurements, which are further combined due to circular nature of the data, as a mixture of von Mises distributions.

3.1 Model of the sound source dynamics

The sound source dynamics is modelled by the well behaved Langevin motion model [12]:

$$\begin{aligned}\begin{bmatrix} \dot{x}_k \\ \dot{y}_k \end{bmatrix} &= \alpha \begin{bmatrix} \dot{x}_{k-1} \\ \dot{y}_{k-1} \end{bmatrix} + \beta \begin{bmatrix} \upsilon_x \\ \upsilon_y \end{bmatrix}, \\ \begin{bmatrix} x_k \\ y_k \end{bmatrix} &= \begin{bmatrix} x_{k-1} \\ y_{k-1} \end{bmatrix} + \delta \begin{bmatrix} \dot{x}_k \\ \dot{y}_k \end{bmatrix},\end{aligned} \quad (7)$$

where where $[x_k, y_k]^{\mathbf{T}}$ is the location of the speaker, $[\dot{x}_k, \dot{y}_k]^{\mathbf{T}}$ is the velocity of the speaker at time index k, $\upsilon_x, \upsilon_y \sim \mathcal{N}(0, \sigma_\upsilon)$ is the stohastic velocity disturbance, α and β are model parameters, and δ is the time between update steps. The system state, i.e. the speaker azimuth, is calculated via the following equation:

$$\mu_k = \operatorname{atan2}\left(\frac{y_k}{x_k}\right). \quad (8)$$

3.2 The von Mises distribution based measurement model

Measurement of the sound source state with M microphones can be described by the following equation:

$$\mathbf{z}_k = \mathbf{h}_k(\mu_k, n_k), \quad (9)$$

where $\mathbf{h}_k(.)$ is a non-linear function with noise term n_k, and $\mathbf{z}_k = [\theta_{ij}^{\pm}, \ldots, \theta_{M,M-1}^{\pm}]_k, i \neq j, \{i,j\} = \{j,i\}$ is the measurement vector defined as a set of azimuths calculated from (6).
Since $\mathbf{z}_k$ is a random variable of circular nature, it is appropriate to model it with the von Mises distribution. The von Mises distribution with its pdf is defined as [15, 16]:

$$p(\theta|\mu,\kappa) = \frac{1}{2\pi I_0(\kappa)} \exp[\kappa\cos(\theta-\mu)], \quad (10)$$

where $0 \leq \theta < 2\pi$ is the measured azimuth, $0 \leq \mu < 2\pi$ is the mean direction, $\kappa > 0$ is the concentration parameter and $I_0(\kappa)$ is the modified Bessel function of the order zero. Bessel function of the order m can be represented by the following infinite sum:

$$I_m(x) = \sum_{k=0}^{\infty} \frac{(-1)^k\,(x)^{2k+|m|}}{2^{2k+|m|}k!\,(|m|+k!)^2}, \quad |m| \neq \frac{1}{2}. \quad (11)$$

Mean direction μ is analogous to the mean of the normal Gaussian distribution, while concentration parameter is analogous to the inverse of the variance in the normal Gaussian distribution. Also, circular variance can be calculated and is defined as:

$$\vartheta^2 = 1 - \frac{I_1(\kappa)^2}{I_0(\kappa)^2}, \quad (12)$$

where $I_1(\kappa)$ is the modified Bessel function of order one.

A microphone pair $\{i,j\}$ at time index k measures two possible azimuths $\theta_{ij,k}^{+}$ and $\theta_{ij,k}^{-}$. Since we cannot discern from a single microphone pair which azimuth is correct, we can say, from a probabilistic point of view, that both angles are equally probable. Therefore, we propose to model each microphone pair as a sum of two von Mises distributions, yielding a bimodal pdf of the following form:

$$\begin{aligned}p_{ij}\left(\theta_{ij,k}^{\pm}|\mu_k,\kappa\right) &= p_{ij}\left(\theta_{ij,k}^{+}|\mu_k,\kappa\right) + p_{ij}\left(\theta_{ij,k}^{-}|\mu_k,\kappa\right) \\ &= \tfrac{1}{2\pi I_0(\kappa)} \exp\left[\kappa\cos\left(\theta_{ij,k}^{+} - \mu_k\right)\right] + \\ &+ \tfrac{1}{2\pi I_0(\kappa)} \exp\left[\kappa\cos\left(\theta_{ij,k}^{-} - \mu_k\right)\right]\end{aligned} \quad (13)$$

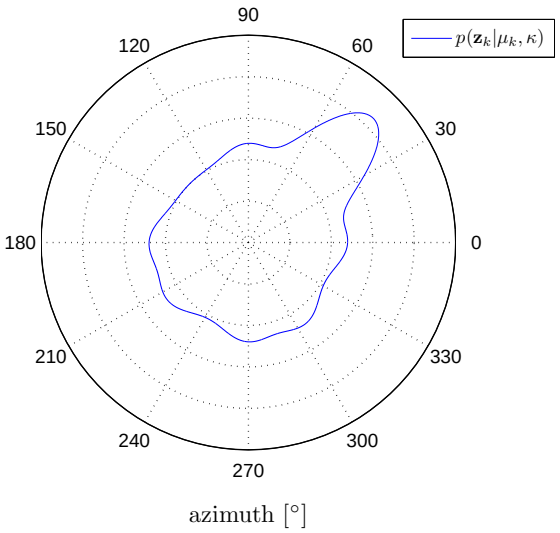


Figure 3: A mixture of several von Mises distributions wrapped on a unit circle (most of them having a mode at $45°$)

Having all pairs modelled as a bimodal distribution, we propose a linear combination of all those pairs to represent the microphone array measurement model. Such a model has the following multimodal pdf:

$$p(\mathbf{z}_k|\mu_k,\kappa) = \frac{1}{2\pi I_0(\kappa)} \sum_{\{i,j\}=1}^{N} \beta_{ij} p_{ij} \left(\theta_{ij,k}^{\pm}|\mu_k,\kappa\right), \tag{14}$$

where N is the total number of microphone pairs and $\sum \beta_{ij} = 1$ is the mixture coefficient.
This model represents our belief in the sound source azimuth. A graphical representation of the analytical (14) is shown in **Figure 3**. Of all the $2N$ measurements, half of them will measure the correct azimuth, while their counterparts from (6) will have different (not equal) values. So, by forming such a linear opinion pool, pdf (14) will have a strong mode at the correct azimuth value.

3.3 Particle filtering

As it was shown, multimodal pdf is inherent to TDOA based localization algorithms and therefore the particle filtering algorithm is utilised, since it is suitable for non-linear systems and measurement equations, non-Gaussian noise and multimodal distributions. This method represents the posterior density function $p(\mu_k|\mathbf{z}_k)$ by a set of random samples (particles) with associated weights and computes estimates based on these samples and weights.
Let $\{\theta_k^p, w_k^p\}_{p=1}^P$ denote a random measure that characterises the posterior pdf $p(\mu_k|\mathbf{z}_k)$, where $\{\theta_k^p,\ p = 1,\dots,P\}$ is a set of support points with associated weights $\{w_k^p,\ p = 1,\dots,P\}$. The weights are normalised so that $\sum_p w_k^p = 1$. Then, the posterior density at k can be approximated as:

$$p(\mu_k|\mathbf{z}_k) \approx \sum_{p=1}^{P} w_k^p \delta(\mu_k - \theta_k^p), \tag{15}$$

where $\delta(.)$ is the Dirac delta measure. Thus, we have a discrete weighted approximation to the true posterior, $p(\mu_k|\mathbf{z}_k)$.
The weights are calculated using the principle of *importance resampling*, where the proposal distribution is given by (7). In accordance to the Sequential Importance Resampling (SIR) scheme, the weight update equation is given by [13]:

$$w_k^p \propto w_{k-1}^p p(\mathbf{z}_k|\mu_k). \tag{16}$$

The next important step in the particle filtering is the *resampling*. The resampling step involves generating a new set of particles by resampling (with replacement) P times from an approximate discrete representation of $p(\mu_k|\mathbf{z}_k)$. After the resampling all the particles have equal weights, which are thus reset to $w_k^p = 1/P$. In the SIR scheme, resampling is applied at each time index. Since we have $w_{k-1}^p = 1/P\ \forall p$, the weights are simply calculated from:

$$w_k^p \propto p(\mathbf{z}_k|\mu_k). \tag{17}$$

The weights given by the proportionality (17) are, of course, normalised before the resampling step. It is also possible to perform particle filter size adaptation through the *KLD-sampling* procedure proposed in [14]. This would take place before the resampling step in order to reduce the computational burden.
To sum up, at each time index k and with M microphones, a set of $2N$ azimuths is calculated with (6), thus forming measurement vector $\mathbf{z}_k$ from which an approximation of (14) is constructed by pointwise evaluation (see **Figure 4**) with particle weights w_k^p calculated from (17).

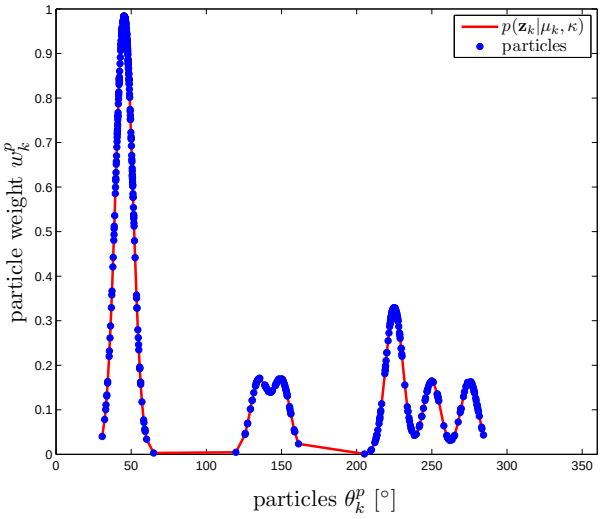
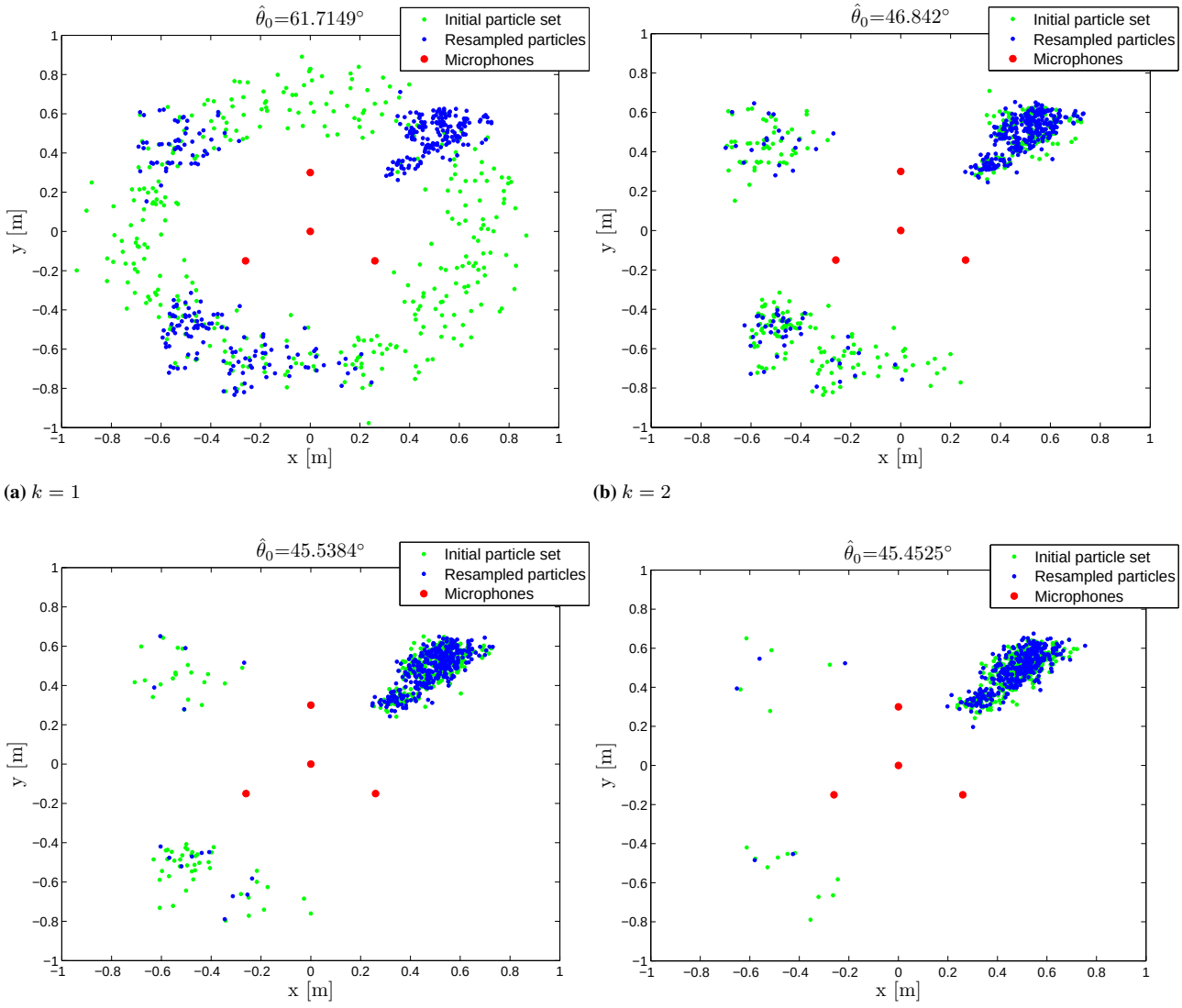


Figure 4: A discrete representation of the true $p(\mathbf{z}_k|\mu_k,\kappa)$ at time index $k = 2$



(a) $k = 1$ **(b)** $k = 2$

(c) $k = 3$ **(d)** $k = 4$

Figure 5: Simulation results

With these particles the azimuth is estimated as follows:

$$\begin{aligned} \hat{\mu}_k &= E\left[\mu_k\right] = \operatorname{atan2}\left(\frac{\mathrm{E}\left[\sin(\theta_k)\right]}{\mathrm{E}\left[\cos(\theta_k)\right]}\right) \\ &= \operatorname{atan2}\left(\frac{\sum_{p=1}^{P} w_k^p \sin(\theta_k^p)}{\sum_{i=p}^{P} w_k^p \cos(\theta_k^p)}\right), \end{aligned} \tag{18}$$

where $\mathrm{E}[.]$ is the expectation operator.

3.4 Algorithm

As it was stated in Section 3, the particle filtering algorithm follows the SIR scheme. The main idea is to spread the particle set $\{\mathbf{s}_k^p, w_k^p\}_{p=1}^P$ in all possible directions, take the measurements $\mathbf{z}_k$, resample the particles with the highest probability, and estimate the azimuth $\hat{\theta}_k$ from their respective weights. After a few steps, most particles will accumulate around the true azimuth value and track the sound source following the motion model given by (7). If at the particular time step k no valid measurements are available (outlier or no voice activity is detected), a Gaussian noise is added to spread the particles to cover a larger area. If this state lasts longer than a given time period, the algorithm is reset and the particles are again spread in all possible directions. **Figure 5** show first four steps of the algorithm execution. The figures show particles before and after the resampling. We can see that the particles converge to the true azimuth value. Detailed description of the algorithm now follows.

Initialization step: At time instant $k = 0$ a particle set $\{s_0^p, w_0^p\}_{p=1}^P$ (velocities $\dot{x}_0, \dot{y}_0$ set to zero) is generated and distributed accordingly on a unit circle. Since the sound

source can be located anywhere around the robot, all the particles have equal weights $w_0^p = 1/P \; \forall p$.

Prediciton step: If there is voice activity detected and the current measurement is valid, all the particles are propagated according to the motion model given by (7). Otherwise, all the particles are corrupted with Gaussian noise. If this state lasts higher than a certain threshold, the algorithm resets to initialization step.

Weight computation: Upon receiving TDOA measurements, DOAs are calculated from (6) and for each DOA a bimodal pdf is constructed from (13). To form the proposed sensor model, all the bimodal pdfs are combined to form (14). The particle weights are calculated from (17) and normalized so that $\sum_{p=1}^{P} w_k^p = 1$.

Azimuth estimation: At this point we have the approximate discrete representation of the posterior density (14). The azimuth is estimated from (18).

Resampling: This step is applied at each time index ensuring that the particles are resampled respective to their weights. After the resampling, all the particles have equal weights: $\{s_k^p, w_k^p\}_{p=1}^P \rightarrow \{s_k^p, 1/P\}_{p=1}^P$. We use the *Systematic resampling* algorithm (see [13]). At this point, before the resampling, it is possible to adapt the particle size by means of *KLD-sampling*. When the resampling is finished, the algorithm loops back to the prediction step.

The algorithm testing was performed with a constructed measurement vector $\mathbf{z}_k$ similar to one that would be experienced during experiments. Six measurements were distributed close to the true value ($\theta = 45°$), while the other six were their counterparts.

4 Experiments

The microphone array used for experiments is composed of 4 omnidirectional microphones arranged in the Y array geometry of side length being $a = 0.5$ cm. The microphone array was placed on a Pioneer 3DX robot as it can bee seen in **Figure 8**. Audio interface is composed of low-cost microphones, pre-amplifiers and external USB soundcard (whole equipment costing cca. €150). Sampling frequency was $F_s = 48$ kHz, 16-bit precision, block length $L = 1024$ samples, and rectangular window was used with zero-padding approach. All the experiments were done in real-time, yielding 21.33 ms system response time.

The first set of experiments was conducted in order to qualitatively asses the performance of the algorithm. In these experiments Y array configuration was used and two scenarios were analyzed. **Figure 6** shows the first scenario, in which a white noise source moved around the mobile robot making a full circle, and the **Figure 7** shows the second scenario, in which a white noise source made rapid angle changes (under 1 second).

Both experiments were repeated with smaller array dimensions (a=30 cm), resulting in smaller angle resolution, and no significant degradations to the algorithm were noticed.

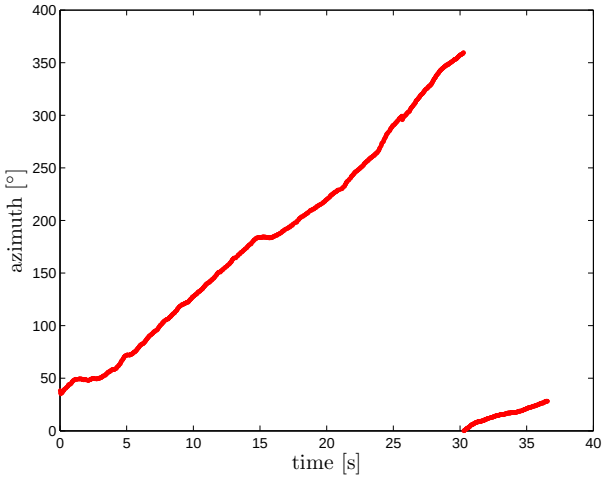


Figure 6: Azimuth estimation for a white noise source making rapid angle changes

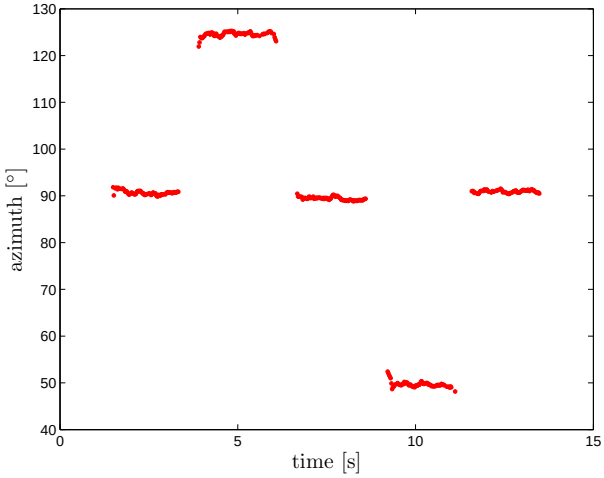


Figure 7: Azimuth estimation for a white noise source making rapid angle changes

5 Conclusion and Future Work

Using a microphone array consisting of 4 omnidirectional microphones, an audio interface for a mobile robot that successfully localizes and tracks a speaker was implemented. The concept is based on a linear combination of probabilistically modelled TDOA measurements. The sensor model uses the proposed von Mises distribution for DOA analysis and for derivation of an adequate azimuth estimation method. In order to handle the inherent multimodal and non-linear characteristics of the system, a particle filtering approach was utilised. However, the proposed post-processing method is not limited to the used sensor.

In order to develop a functional human-aware mobile robot system, future works will strive towards the integration of the proposed algorithm with other systems like leg tracking, robot vision etc. Also, by utilising a TDOA estima-

tion method that is capable of tracking multiple speakers, further capabilities of the proposed sensor model could be researched.

Acknowledgment

This work was supported by the Ministry of Science, Education and Sports of the Republic of Croatia under grant No. 036-0363078-3018.

Figure 8: The robot and the microphone array used in the experiments

References

[1] Valin J.-M.; Michaud F.; Rouat J.: *Robust Localization and Tracking of Simultaneous Moving Sound Sources Using Beamforming and Particle Filtering*, Robotics and Autonomous Systems, vol. 55, no. 3, pp. 216–228, 2007.

[2] DiBiase J. H.; Silverman H. F.; Brandstein M. S.: *Robust Localization in Reverberant Rooms*, in Microphone Arrays: Signal Processing Techniques and Applications, Springer, 2001.

[3] Di Caludio E. D.; Parisi R.: *Multi Source Localization Strategies*, in Microphone Arrays: Signal Processing Techniques and Applications, Springer, 2001.

[4] Merimaa J.: *Analysis, Synthesis, and Perception of Spatial Sound - Binaural Localization Modeling and Multichannel Loudspeaker Reproduction*. PhD thesis, Helsinki University of Technology, 2006.

[5] Ferreira J. F.; Pinho C.; Dias J.: *Implementation and Calibration of a Bayesian Binaural System for 3D Localisation*, in Proceedings of the 2008 IEEE International Conference on Robotics and Biomimetics, pp. 1722–1727, 2009.

[6] Marković I.; Petrović I.: *Speaker Localization and Tracking in Mobile Robot Environment Using a Microphone Array*, in Proceedings book of 40th International Symposium on Robotics (J. Basañez, Luis; Suárez, Raúl ; Rosell, ed.), no. 2, (Barcelona), pp. 283–288, Asociación Española de Robótica y Automatización Tecnologías de la Producción, 2009.

[7] Nishiura T.; Nakamura M.; Lee A.; Saruwatari H.; Shikano K.: *Talker Tracking Display on Autonomous Mobile Robot with a Moving Microphone Array*, in Proceedings of the 2002 International Conference on Auditory Display, pp. 1–4, 2002.

[8] Brandstein M.; Ward D.: *Microphone Arrays: Signal Processing Techniques and Applications*. Springer, 2001.

[9] Chen J.; Benesty J.; Huang Y. A.: *Time Delay Estimation in Room Acoustic Environments: an overview*, EURASIP Journal on Applied Signal Processing, pp. 1–19, 2006.

[10] Knapp C.; Carter G.: *The Generalized Correlation Method for Estimation of Time Delay*," IEEE Transactions on Acoustics Speech and Signal Processing, vol. 24, no. 4, p. 320âĂŞ327, 1976.

[11] Donohue K. D.; Hannemann J.; and Dietz H. G.: *Performance of Phase Transform for Detecting Sound Sources with Microphone Arrays in Reverberant and Noisy Environments*, Signal Processing, vol. 87, pp. 1677–1691, 2007.

[12] Vermaak J.; Blake A.: *Nonlinear Filtering for Speaker Tracking in Noisy and Reverberant Environments*, in Proceeding of the IEEE International Conference on Acoustic, Speech and Signal Processing, 2001.

[13] Arulampalam S.; Maskell S.; Gordon N.; Clapp T.: *A Tutorial on Particle Filters for On-line Nonlinear/Non-Gaussian Bayesian Tracking*, Signal Processing, vol. 50, pp. 174–188, 2001.

[14] Fox D.: *Adapting the Sample Size in Particle Filter Through KLD-Sampling*, International Journal of Robotics Research, vol. 22, 2003.

[15] Mardia K. V.; Jupp P. E.: *Directional Statistics*. New York: Wiley, 1999.

[16] Fisher N. I.: *Statistical Analysis of Circular Data*. Cambridge University Press, 1996.