Interspeech 2018
2-6 September 2018, Hyderabad

Multiple Concurrent Sound Source Tracking Based on Observation-Guided Adaptive Particle Filter

Hong Liu, Haipeng Lan, Bing Yang, Cheng Pang

Key Laboratory of Machine Perception, Peking University, Shenzhen Graduate School.
{hongliu, haipenglan}@pku.edu.cn, {bingyang, chengpang}@sz.pku.edu.cn

Abstract

Particle filter (PF) has been proved to be an effective tool to track sound sources. In traditional PF, a pre-defined dynamic model is used to model source motion, which tends to be mismatched due to the uncertainty of source motion. Besides, non-stationary interferences pose a severe challenge to source tracking. To this end, an observation-guided adaptive particle filter (OAPF) is proposed for multiple concurrent sound source tracking. Firstly, sensor signals are processed in the time-frequency domain to obtain the direction of arrival (DOA) observations of sources. Then, by updating particle states with these DOA observations, angular distances between particles and observations are reduced to guide particles to directions of sources. Thirdly, particle weights are updated by an interference-adaptive likelihood function to reduce the impacts of interferences. At last, with the updated particles and the corresponding weights, OAPF is utilized to determine the final DOAs of sources. Experimental results demonstrate that our method achieves favorable performance for multiple concurrent sound source tracking in noisy environments.

Index Terms: multiple concurrent sound source tracking, particle filter, direction of arrival

1. Introduction

Multiple concurrent sound source tracking using microphone arrays plays a crucial role in numerous applications such as speech enhancement, teleconferencing and human-robot interaction. However, the performance of source tracking is affected by the uncertainty of source motion and the non-stationary of interference. A method for the robust sound source tracking in such complex conditions is demanded.

Over the past few decades, many sound source tracking methods have been proposed, which generally consist of two stages: 1) Localization stage. The current received signals are converted to position observations of sound sources by using localization function. The position observation can be time difference of arrival [1, 2], location information [3], direction of arrival (DOA) [4, 5] or binaural cues [6–8]. 2) Tracking stage. When the position observations across successive frames are obtained, to determine the final source position estimations, the temporal consistency of position observations is usually exploited by a filter method, such as Kalman filter (KF) [9], extended Kalman filter (EKF) [10,11] and particle filter (PF) [12–18]. Different from KF and EKF which are based on the assumption of linearity or Gaussianity, PF can overcome the limitation of the assumption, which has been proved to be a promising method in tracking problems. It estimates source positions by a set of particles with associated weights.

Generally, in PF, a pre-defined dynamic model is assumed to be able to model the source motion effectively in the prediction step, making it possible to predict the propagation of particle states recursively according to the dynamic model. Due to the uncertainty of source motion, the dynamic model may mismatch the source motion. The propagation of particles may be misled, causing particles away from sources. Hence, the EKF is introduced into the extended Kalman particle filter (EKPF) [17], which can use the latest observations to update particle states. EKPF is based on the Gaussianity assumption which may be invalid in fact, resulting in tracking losses. Swarm intelligence based particle filter (SWIPF) [18] updates particle states by using particle swarm optimization algorithm [19]. Since it employs the interaction between particles instead of the latest observations, particles are guided to source positions in an indirect way, resulting in tracking performance degradation.

In addition, the likelihood function, which depends on observation interference distribution [20], is exploited to update particle weights in the update step. To track sources simultaneously, a bank of parallel PFs is used [13]. In this case, each PF needs to obtain the corresponding likelihood function. Generally, the distribution of the interferences exerted on different observations is assumed to be known and same, which can be modelled as Gaussian distribution [13] and Von Mises distribution [15]. Since source observations are not only affected by noises but also by the interaction between sources, the interferences exerted on different observations may vary with time and environments. Thus, the distribution of the interferences may be uncertain and different due to the non-stationary of interferences. The pseudo likelihood function, which is determined by the localization function, is exploited [14]. However, the pseudo likelihood function reflects the joint probability distribution of observations, it cannot be applied to the PF bank directly.

To deal with the problems mentioned above, a novel multiple concurrent sound source tracking method based on observation-guided adaptive particle filter (OAPF) is proposed. Firstly, received signals are processed in the time-frequency (T-F) domain to obtain the direction of arrival (DOA) observations of sources. Secondly, DOA observations are used to update the directions of particles to reduce the angular distances between sources and particles. The updated particle directions are exploited to modify the states of the corresponding particles. In this way, OAPF can directly guide particles to source directions by current observations. Thirdly, since the interferences impacted on DOA observations can affect the half-width at half-maximum (HWHM) [21] of peaks of sources in the localization function, an interference-adaptive likelihood function is proposed which takes HWHM as its parameter. Therefore, the adaptive likelihood function can be adaptively adjusted according to the influence of non-stationary interferences. Then, the likelihood function is used to update particle weights, and the particles with small weights are abandoned in resample step. Therefore, OAPF can adaptively adjust particle distribution according to the interference influence. Finally, OAPF uses the particles and their weights to estimate the DOAs of sources.

10.21437/Interspeech.2018-1248

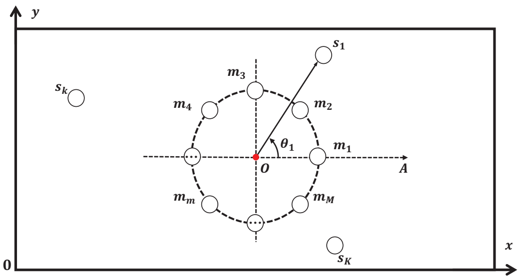

Figure 1: *The illustration of the geometrical relationship. Here, an equispaced circular sensor array consists of M microphones and K sources move around the array.*

2. Multiple Concurrent Sound Source Localization

There are K far-field sound sources observed by a M-channel array as depicted in Figure 1, the sources move simultaneously while the array is static. Here, the DOAs of sources are adopted as position observations, which are observed with respect to the ray OA.

To obtain the DOA observations, a sparse component analysis based multiple concurrent sound source localization method is adopted [4]. The number of sources is assumed to be known. The sensor signals are converted into TF domain by the short-time Fourier transform firstly. Since speech signals are sparse in TF domain, there exist some TF bins where one source is dominant over others in energy, and these bins are known as single source bins. Then, the coarse DOAs of sources can be estimated independently in every single source bin, which are used to construct a DOA distribution histogram. Thirdly, the histogram is smoothed by Parzen windows [22] and the smoothed histogram is adopted as the localization function. Finally, the DOA estimations are refined by applying matching pursuit method [23] on the smoothed histogram, and the refined DOAs are adopted as observations. Let $\boldsymbol{X}_t$ denote the signals received by the sensor array at time t, the localization function can be formulated:

$$\boldsymbol{Y}(\hat{\boldsymbol{\theta}}_t) = F_T\{\boldsymbol{X}_t\}, \tag{1}$$

where $\hat{\boldsymbol{\theta}}_t = [\hat{\theta}_{1,t}, ..., \hat{\theta}_{k,t}, ..., \hat{\theta}_{K,t}]$ is the DOA observations of sources, $\boldsymbol{Y}(\hat{\boldsymbol{\theta}}_t)$ is the localization function and $F_T\{\cdot\}$ is the function which transforms $\boldsymbol{X}_t$ into DOA information.

3. Multiple Concurrent Sound Source Tracking Based on OAPF

To deal with the mismatch problem and the non-stationary interference problem, an observation-guided adaptive particle filter (OAPF) is proposed. In OAPF, the latest observations are used to modify particle states and an interference-adaptive likelihood function is employed to update particle weights. To track sources simultaneously, an OAPF bank is employed which is denoted as $\{\mathcal{F}_{k,t}\}$ with $k \in \{1, ..., K\}$. Each OAPF is equipped with Z particles, which is denoted as $\{\boldsymbol{\alpha}^z_{k,t}, w^z_{k,t}\}$ with $z \in \{1, ..., Z\}$. The particle state is defined as:

$$\boldsymbol{\alpha}^z_{k,t} = \left[\mathcal{X}^z_{k,t}, \dot{\mathcal{X}}^z_{k,t}, \mathcal{Y}^z_{k,t}, \dot{\mathcal{Y}}^z_{k,t}\right]^\mathsf{T}, \tag{2}$$

where $\mathcal{X}^z_{k,t}$ and $\mathcal{Y}^z_{k,t}$ are the coordinates of the particle, $\dot{\mathcal{X}}^z_{k,t}$ and $\dot{\mathcal{Y}}^z_{k,t}$ are the velocities, and $[\cdot]^\mathsf{T}$ denotes the transpose. Here only a two-dimensional tracking situation is considered. Initially, particles of an OAPF is drawn from the importance density [12] with same weight $\{w^z_{k,0} = \frac{1}{Z}\}$. Then the OAPF is employed to track the corresponding source using a two-step process of prediction and update.

3.1. Particle State Propagation Based on DOA Guidance

In OAPF, the dynamic model and the current source observations are jointly exploited for the propagation of particles. Here the Langevin model is adopted as the dynamic model. In X-coordinate, it is defined as:

$$\dot{\mathcal{X}}_t = \dot{\mathcal{X}}_{t-1} \cdot e^{-\gamma\Delta T} + \Gamma_{\mathcal{X}} \cdot \varsigma \cdot \sqrt{1 - e^{-2\gamma\Delta T}}, \tag{3}$$

$$\mathcal{X}_t = \mathcal{X}_{t-1} + \Delta T \cdot \dot{\mathcal{X}}_t, \tag{4}$$

where $\gamma = 10\ s^{-1}, \varsigma = 1\ ms^{-1}$, ΔT is the time interval between two successive frames and $\Gamma_{\mathcal{X}}$ is a normally distributed random variable. Parameters of the Langevin model in Y-coordinate are identical. By using the dynamic model, particle states can be predicted recursively.

Then, the coordinate information of the predicted particles is converted to direction information as:

$$\tilde{\theta}^z_{k,t} = arctan(\frac{\mathcal{Y}^z_{k,t} - \mathcal{Y}'}{\mathcal{X}^z_{k,t} - \mathcal{X}'}), \tag{5}$$

where $\tilde{\theta}^z_{k,t}$ is the direction of the predicted particle and $(\mathcal{X}', \mathcal{Y}')$ represents the coordinates of the array center.

Given the DOA observations, directions of particles are updated to reduce angular distances between sources and particles:

$$\bar{\theta}^z_{k,t} = \tilde{\theta}^z_{k,t} + \beta(\hat{\theta}_{k,t} - \tilde{\theta}^z_{k,t}), \tag{6}$$

where $\hat{\theta}_{k,t}$ is the observation of source k, $\bar{\theta}^z_{k,t}$ is the updated direction of the corresponding particle, β is an updating parameter. A larger β indicates that the particles are directed more closely to the source direction. Here, $\beta = 0.28$.

Then, the directions of the updated particles are exploited to modify the coordinates of the corresponding particles:

$$\left[\bar{\mathcal{X}}^z_{k,t}, \bar{\mathcal{Y}}^z_{k,t}\right]^\mathsf{T} = d^z_{k,t} \cdot \left[cos(\bar{\theta}^z_{k,t}), sin(\bar{\theta}^z_{k,t})\right]^\mathsf{T} + \left[\mathcal{X}', \mathcal{Y}'\right]^\mathsf{T}, \tag{7}$$

where $d^z_{k,t} = \sqrt{(\mathcal{X}^z_{k,t} - \mathcal{X}')^2 + (\mathcal{Y}^z_{k,t} - \mathcal{Y}')^2}$. In this way, particle states can be guided by the observations, thereby improving the convergence to the source direction.

Therefore, the modified state of the particle is:

$$\boldsymbol{\alpha}^z_{k,t} = \left[\bar{\mathcal{X}}^z_{k,t}, \dot{\mathcal{X}}^z_{k,t}, \bar{\mathcal{Y}}^z_{k,t}, \dot{\mathcal{Y}}^z_{k,t}\right]^\mathsf{T}. \tag{8}$$

3.2. Adaptive Likelihood Function Calculation

In the localization stage, the coarse DOAs are estimated from single source bins. When the interferences exert slighter influences on source observations, the angular distances between source DOAs and the coarse DOAs associated with sources will decrease. Besides, the amount of the coarse DOAs resulting by interferences decreases. Hence, the amount of the coarse DOAs associated with sources increases relatively and they are more concentrated around the source positions. Correspondingly, the widths of the corresponding source peaks tend to be narrower in the localization function $\boldsymbol{Y}(\hat{\boldsymbol{\theta}}_t)$. Here, the half-width at half-maximum (HWHM) [21] is employed to measure the width information of peaks. Figure 2 illustrates the HWHMs of peaks in the localization function under different noisy environments. It can be observed that the HWHM increases when SNR decreases. Moreover, the HWHMs of two sources are different although the sources are under the same environment. It confirms that different observations suffer different interference influences. Therefore, distribution of the interferences exerted on different observations are different.

Motivated by the relation between the HWHM of source peak and the interference influence, a Cauchy probability den-

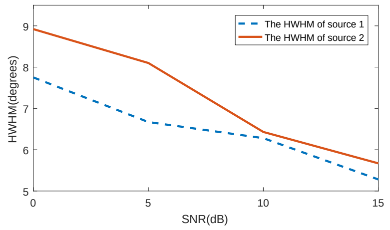

Figure 2: *The HWHM of sources in two-source environments for different noise levels. There are two active sources in the same environment. The distances from sources to array center are* 1.5*m.*

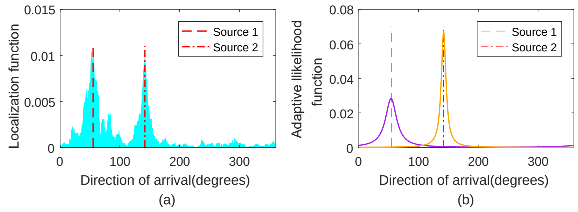

Figure 3: *A two-source instance in a noisy room where sources are located at* 52° *and* 142° *respectively. (a) The localization function. (b) The adaptive likelihood functions.*

sity [21] based adaptive likelihood function is proposed, which takes HWHM as its parameter. Different from the DOA distribution which is defined in circular axis, the Cauchy probability density is defined in linear axis. Hence, by converting the Cauchy probability density into a wrapped form, the adaptive likelihood function is proposed as:

$$\mathbb{C}(\theta;\hat{\theta}_{k,t},\psi_{k,t},\rho)=\frac{1}{\pi}\left[\frac{\psi_{k,t}\cdot\rho}{\eta\cdot cos(\theta-\hat{\theta}_{k,t})^3+(\psi_{k,t}\cdot\rho)^2}\right], \quad (9)$$

where $\psi_{k,t}$ is the HWHM corresponding to the DOA observation $\hat{\theta}_{k,t}$, η is used to adjust the peak value of the adaptive likelihood function and ρ is exploited to scale HWHMs. Here, $\eta = 3200$ and $\rho = 0.5$. The likelihood function represents the probability distribution of DOA observation, which is determined by the DOA observation and the corresponding observation interference distribution. Different from traditional likelihood functions, the adaptive likelihood considers the difference of the influences on different sources, which employs the HWHM to reflect the observation interference distribution.

A two-source instance is depicted in Figure 3. Although the heights of two source peaks are almost the same, the HWHMs of two peaks are different. The difference indicates that the observations of the sources suffer interference influence in different degrees. A smaller HWHM leads to a smaller dispersion of observation probability distribution.

3.3. DOA Estimation Based on Adaptive Likelihood Function

Given the DOA observations, the corresponding adaptive likelihood functions can be determined, which are used to update particle weights. According to the adaptive likelihood function, the probability of attaining observation $\hat{\theta}_{k,t}$ conditioned on a given particle state $\alpha^z_{k,t}$ is:

$$p(\hat{\theta}_{k,t}|\alpha^z_{k,t})=\mathbb{C}(DOA(\alpha^z_{k,t});\hat{\theta}_{k,t},\psi_{k,t},\rho), \quad (10)$$

where $DOA(\alpha^z_{k,t})$ is the particle direction, which can be calculated by Eq. (5).

The weight of the corresponding particle can be updated:

$$w^z_{k,t}=w^z_{k,t-1}\cdot p(\hat{\theta}_{k,t}|\alpha^z_{k,t}). \quad (11)$$

Using the particles and their updated weights, the source DOA can be obtained:

$$\theta_{k,t}=\sum_{z=1}^{Z}\frac{w^z_{k,t}}{\sum_{z=1}^{Z}w^z_{k,t}}DOA(\alpha^z_{k,t}). \quad (12)$$

Algorithm 1: Resample algorithm in particle filter

Input: $\{\alpha^{(1:Z)}, w^{(1:Z)}\}$
Output: $\{\alpha^{*(1:Z)}, w^{(1:Z)}\}$
Set the cumulative distribution function (CDF): $C_1 = 0$
for z= $2:Z$ **do**
  $C_z = C_{z-1} + w^z$ $\rightarrow$ Construct CDF
end for
Set $z = 1$ $\rightarrow$ Start at the bottom of the CDF
Draw u_1 from the uniform distribution $U[0, 1/Z]$
for $h = 1 : Z$ **do**
  $u_h = u_1 + (h-1)/Z$ $\rightarrow$ Move along the CDF
  while $u_h > C_z$ **do**
    $z = z + 1$
  end while
  $\alpha^{*h} = \alpha^z$,$w^h = 1/Z$ $\rightarrow$ Assign particle and weight
end for
return $\{\alpha^{*(1:Z)}, w^{(1:Z)}\}$

A resample algorithm [24] is utilized to alleviate the degeneracy of particles. The details of resample are shown in Algorithm 1. Particles will be resampled if the particle set satisfies:

$$\frac{1}{\sum_{z=1}^{Z}(w^z_{k,t})^2}<N_{thr}, \quad (13)$$

where N_{thr} is a pre-defined threshold and $N_{thr} = Z$. In this step, particles with relatively smaller weights are eliminated while those with larger weights are duplicated. The resample of particles indicates that a larger HWHM corresponds a broader particle distribution. In this way, OAPF can adjust particle distribution adaptively according to the influence exerted on the observations.

Since the reliability of a DOA observation decreases when the interference influence becomes severer, the estimation of DOA observation tends to have larger error. Due to the misestimated DOA observation, particles may be sampled from an erroneous region. In this case, OAPF is allowed to sample particles in a wider region through a larger HWHM to strengthen the robustness to interferences.

Considering the temporal consistency of the observations associated with the sources, a correlation measurement is proposed to assign the observations to the corresponding source. The correlation measurement between source k and a DOA observation is defined:

$$W_{k,k'}=\int_0^{2\pi}\mathbb{C}(\theta;\theta_{k,t-1},\psi_{k,t-1},\rho)\cdot\mathbb{C}(\theta;\hat{\theta}_{k',t},\psi_{k',t},\rho)d\theta. \quad (14)$$

Each source calculates the correlation measurements with all DOA observations, a DOA observation is assigned to the source which has the highest correlation measurement value with this DOA observation.

4. Experiment and Discussions

4.1. Experimental Setting

Performances of proposed method are evaluated in a noisy room with the size of 6 m × 4 m × 3 m. An 8-channel uniform circular array with a radius of 2 cm is placed at the center of the rectangular room. The coordinates of the microphone m_1

Table 1: *The average RMSEs (in degrees) of tracking methods based on different particle update strategies*

SNR	methods	30°	60°	90°	135°
15*dB*	**proposed**	**1.20**	**1.36**	**1.26**	**1.24**
	SWIPF	1.74	1.90	1.84	1.74
	SIRPF	1.92	1.82	1.93	1.63
10*dB*	**proposed**	**1.66**	**1.70**	**1.65**	**1.50**
	SWIPF	2.86	2.62	2.63	2.58
	SIRPF	3.04	2.88	2.95	2.52
5*dB*	**proposed**	**2.51**	**2.73**	**2.36**	**2.16**
	SWIPF	3.83	3.71	3.35	3.25
	SIRPF	4.33	4.83	4.57	3.82
0*dB*	**proposed**	**3.85**	**4.77**	**3.97**	**3.59**
	SWIPF	5.79	5.81	5.09	4.99
	SIRPF	6.43	7.59	7.85	6.10

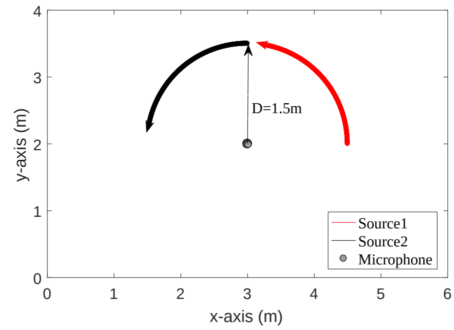

Figure 4: *The trajectories of two sources in the simulated environment. The two sources move around the array with a fixed velocity. Here the angular separation of two sources is* 90° .

is $(3.02m, 2m, 1.7m)$. To generate multiple concurrent source scenes, two sources are located around the array with 1.5 m distance. The sources with a height of 1.7 m move anticlockwise with the same speed of 0.5 m/s as Figure 4 shows. The angular separations between sound sources are set to 30°, 60°, 90° and 135°. Ten pairs of speech recordings from TSP Speech Database [25] are employed as the source signals. The sampling rates of the recordings are 48 kHz. The ISM toolbox [26] is employed to generate the room impulse responds from sources to the microphones. To obtain an objective signal-to-noise ratio (SNR), the white Gaussian noise is properly scaled and added to each microphone signal. The frame length is set to 2048 with 50% overlap, and the highest frequency of interest is set to 4 kHz. The particle number Z is 100. The average Root Mean Square Error (RMSE) is adopted as evaluation criteria [14].

4.2. Experimental Results

To test the effectiveness of the DOA guidance, the proposed method is compared with tradition PF [13] and SWIPF [18]. The tradition PF is known as sequential importance resampling particle filter (SIRPF). The three methods adopt the adaptive likelihood function to update particle weights in the update step. Table I illustrates the comparison of RMSE for the methods in different noisy environments. It can be seen that our method outperforms SWIPF and SIRPF under all conditions. All the RMSEs of our method are less than 5° while the maximum RMSEs of the other two methods are larger than 5°. Since OAPF employs source observations to reduce angular distances between sources and particles, the particles can be more closely to the source direction. Thus particles can achieve a more accurate

Table 2: *The average RMSEs (in degrees) of tracking methods based on different likelihood functions.*

SNR	methods	30°	60°	90°	135°
15*dB*	**proposed**	**1.20**	**1.36**	**1.26**	**1.24**
	VMPF	2.55	2.42	2.39	2.48
10*dB*	**proposed**	**1.66**	**1.70**	**1.65**	**1.50**
	VMPF	3.23	2.99	3.22	2.92
5*dB*	**proposed**	**2.51**	**2.73**	**2.36**	**2.16**
	VMPF	3.72	3.75	3.40	3.47
0*dB*	**proposed**	**3.85**	**4.77**	**3.97**	**3.59**
	VMPF	5.77	4.99	5.88	4.99

approximation. Different from SWIPF, our method takes use of source observations instead of inter-particle interaction information to update particle states. Hence, the particles are guided to the sources in a more direct way to improve the tracking performance. The performance of all the three methods tends to be worse when the noise level is relatively larger.

To demonstrate the effectiveness of the adaptive likelihood function, our method is compared with the Von Mises likelihood based particle filter (labeled as VMPF). In the VMPF, the Von Mises distribution, which is a circular distribution, is exploited to model the observation interference distribution. Same as the proposed method, the VMPF also exploits observations to modify particle states in prediction stage. Table II depicts the comparison of RMSE for the methods based on different likelihood functions. It is observed that the proposed method performs better than VMPF in all the cases. The RMSE differences between the VMPF and our method are larger than 1° in most cases. This is due to that the distribution of particles can be adaptively modified according to the interference influence. When sources suffer slighter influence, the observations are more reliable. In this case, OAPF samples particles from the region with high likelihood to achieve a more accurate estimation. When the reliability of observation decreases, OAPF samples particles from a border region to avoid the observation misdirection. In this way, OAPF shows the adaptivity to interferences and strengthens the robustness to interferences. In summary, compared with other methods, the proposed method is more reliable and practical for source tracking.

5. Conclusions

This paper proposes an effective multiple concurrent sound source tracking method based on OAPF for noisy environments. By reducing angular distances between particles and the corresponding sources observation, OAPF can exploit the observations to guide particles towards source directions when the dynamic model can not model the source motion properly. The adaptive likelihood function can be adjusted according to the interferences exerted on sources by using HWHM, and it is used to update the weights of particles. Therefore, OAPF can adaptively modify the particles distribution to strengthen robustness to the interferences. Experimental results show that the proposed method can track multiple concurrent sound sources with high accuracy. Future work will concentrate on the performance of tracking method in reverberant environments.

6. Acknowledgements

This work is supported by National Natural Science Foundation of China (NSFC No.61673030, U1613209), Natural Science Foundation of Guangdong Province (No. 2015A030311034).

7. References

[1] C. Knapp and G. Carter, "The generalized correlation method for estimation of time delay," *IEEE Transactions on Acoustics Speech and Signal Processing*, vol. 24, no. 4, pp. 320–327, 1976.

[2] K. Cho, T. Nishiura, and Y. Yamashita, "A study on multiple sound source localization with a distributed microphone system," in *INTERSPEECH*, 2009, pp. 1359–1362.

[3] M. F. Fallon and S. Godsill, "Acoustic source localization and tracking using track before detect," *IEEE Transactions on Audio Speech and Language Processing*, vol. 18, no. 6, pp. 1228–1242, 2010.

[4] D. Pavlidi, A. Griffin, M. Puigt, and A. Mouchtaris, "Real-time multiple sound source localization and counting using a circular microphone array," *IEEE Transactions on Audio Speech and Language Processing*, vol. 21, no. 10, pp. 2193–2206, 2013.

[5] B. Yang, H. Liu, and C. Pang, "Multiple sound source counting and localization based on spatial principal eigenvector," in *INTERSPEECH*, 2017, pp. 1924–1928.

[6] J. C. Middlebrooks and D. M. Green, "Sound localization by human listeners." *Annual Review of Psychology*, vol. 42, no. 1, pp. 135–159, 1991.

[7] Z. Jie and L. Hong, "Robust acoustic localization via time-delay compensation and interaural matching filter," *IEEE Transactions on Signal Processing*, vol. 63, no. 18, pp. 4771–4783, 2015.

[8] C. Pang, H. Liu, J. Zhang, and X. Li, "Binaural sound localization based on reverberation weighting and generalized parametric mapping," *IEEE/ACM Transactions on Audio Speech and Language Processing*, vol. 25, no. 8, pp. 1618–1632, 2017.

[9] Y. Oualil, F. Faubel, and D. Klakow, "A multiple hypothesis gaussian mixture filter for acoustic source localization and tracking," in *International Workshop on Acoustic Signal Enhancement*, 2012, pp. 1–4.

[10] S. Gannot and T. G. Dvorkind, "Microphone array speaker localizers using spatial-temporal information," *Eurasip Journal on Advances in Signal Processing*, vol. 2006, no. 1, pp. 1–17, 2006.

[11] Y. Oualil and D. Klakow, "Multiple concurrent speaker short-term tracking using a kalman filter bank," in *IEEE International Conference on Acoustics, Speech and Signal Processing*, 2014, pp. 1458–1462.

[12] J. Vermaak and A. Blake, "Nonlinear filtering for speaker tracking in noisy and reverberant environments," in *IEEE International Conference on Acoustics, Speech, and Signal Processing*, 2001, pp. 3021–3024.

[13] J. M. Valin, F. Michaud, and J. Rouat, "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.

[14] 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*, vol. 11, no. 6, pp. 826–836, 2003.

[15] I. Marković and I. Petrović, "Speaker localization and tracking with a microphone array on a mobile robot using von mises distribution and particle filtering," *Robotics and Autonomous Systems*, vol. 58, no. 11, pp. 1185–1196, 2010.

[16] W. K. Ma, B. N. Vo, S. S. Singh, and A. Baddeley, "Tracking an unknown time-varying number of speakers using tdoa measurements: a random finite set approach," *IEEE Transactions on Signal Processing*, vol. 54, no. 9, pp. 3291–3304, 2006.

[17] X. Zhong and J. R. Hopgood, "Particle filtering for tdoa based acoustic source tracking: Nonconcurrent multiple talkers," *Signal Processing*, vol. 96, no. 5, pp. 382–394, 2014.

[18] K. Wu, A. W. H. Khong, and T. G. Shu, "Swarm intelligence based particle filter for alternating talker localization and tracking using microphone arrays," *IEEE/ACM Transactions on Audio Speech and Language Processing*, vol. 25, no. 6, pp. 1384–1397, 2017.

[19] R. Poli, J. Kennedy, and T. Blackwell, "Particle swarm optimization," *Swarm Intelligence*, vol. 1, no. 1, pp. 33–57, 2007.

[20] A. Mukherjee and A. Sengupta, "Likelihood function modeling of particle filter in presence of non-stationary non-gaussian measurement noise," *Signal Processing*, vol. 90, no. 6, pp. 1873–1885, 2010.

[21] N. L. Johnson, S. Kotz, and N. Balakrishnan, *Continuous Univariate Distributions, Volume 1, 2nd Edition*. Wiley, 1994, ch. 16.

[22] R. O. Duda, P. E. Hart, and D. G. Stork, *Pattern Classification (2nd Edition)*. Wiley, 2000.

[23] S. G. Mallat and Z. Zhang, "Matching pursuits with time-frequency dictionaries," *IEEE Transactions on Signal Processing*, vol. 41, no. 12, pp. 3397 – 3415, 1993.

[24] M. S. Arulampalam, S. Maskell, N. Gordon, and T. Clapp, "A tutorial on particle filters for online nonlinear/non-gaussian bayesian tracking," *IEEE Transactions on Signal Processing*, vol. 50, no. 2, pp. 174–188, 2002.

[25] P. Kabal, *"TSP Speech Database"*, Department of Electrical and Computer Engineering, McGill University, Montreal Canada, 2002.

[26] E. A. Lehmann and A. M. Johansson, "Diffuse reverberation model for efficient image-source simulation of room impulse responses," *IEEE Transactions on Audio Speech and Language Processing*, vol. 18, no. 6, pp. 1429–1439, 2010.