IDIAP RESEARCH REPORT

A TDOA GAUSSIAN MIXTURE MODEL FOR IMPROVING ACOUSTIC SOURCE TRACKING

Youssef Oualil Friedrich Faubel
Mathew Magimai.-Doss Dietrich Klakow

Idiap-RR-10-2012

MARCH 2012

Centre du Parc, Rue Marconi 19, P.O. Box 592, CH - 1920 Martigny
T +41 27 721 77 11 F +41 27 721 77 12 info@idiap.ch www.idiap.ch

A TDOA GAUSSIAN MIXTURE MODEL FOR IMPROVING ACOUSTIC SOURCE TRACKING

Youssef Oualil[1,2]*, Friedrich Faubel*[1]*, Mathew Magimai Doss*[2]*, Dietrich Klakow*[1]

[1] Spoken Language Systems, Saarland University, Saarbrücken, Germany
[2] Idiap Research Institute, Martigny, Switzerland
{youssef.oualil,friedrich.faubel}@lsv.uni-saarland.de

ABSTRACT

Traditionally, *time difference of arrival* (TDOA) based acoustic source tracking consists of two stages, more precisely, estimation of TDOAs followed by a tracking algorithm. In general, these two stages are performed separately and presume that (1) TDOAs can be estimated reliably; and (2) the errors in detection behave in a well-defined fashion. The presence of noise and reverberation, however, leads to multimodal TDOA distributions and causes larger errors in the estimates, which ultimately lowers the tracking performance. To counteract this effect, we propose an approach that enhances TDOA estimation by (1) accounting for the multimodal aspect through a Gaussian mixture model and (2) integrating knowledge that has been obtained in the tracking stage. In doing so, this approach tightly couples the two stages. Experimental results on the AV16.3 corpus show that the proposed approach improves the tracking performance significantly compared to various other tracking algorithms.

Index Terms— Direction of arrival estimation, Tracking, Microphone Arrays, Kalman filters

1 Introduction

Tracking acoustic sources is becoming, increasingly, more important, with the increase in number of applications, such as (multiparty) speech enhancement/separation, automatic camera steering, etc. TDOA-based source tracking solves this problem in two stages, namely a detection stage and a tracking stage. In the detection stage, the TDOA which is introduced at each sensor pair is estimated, typically, under use of the generalized cross correlation (GCC) [1]. In the tracking stage, the source position is triangulated in a consistent fashion by integrating the estimated TDOAs through use of a Kalman filter extension or a particle filter [2, 3, 4]. Unfortunately, the tracking performance degrades due to noise and multi-path effects. For instance, under room acoustical conditions, early reflections and reverberation corrupt the GCCs through smearing and through introduction of secondary peaks [5, 6]. This in turn affects the tracking algorithms, which assume the error is a stationary Gaussian process whereas the TDOA error in a multi-path environment is rather time-varying and multimodal.

Motivated by previous works [4, 6], this paper proposes a novel probabilistic approach, which enhances TDOA estimates by interpreting the normalized GCC as a probability density function (pdf) of the TDOAs. More precisely in this approach, (1) a Gaussian distribution is associated to each GCC peak, as a consequence of which the TDOA pdf is approximated by a Gaussian mixture model (GMM). Such an approximation is realistic because it takes into account the multimodal aspect of TDOAs. In addition, it also allows us to integrate knowledge that has been obtained in the tracking stage. Then, (2) the TDOA pdf, which the tracking algorithm expects at the current time instant, is *predicted* and the mixture weights of the above GMM are updated by measuring the "similarity" between each of its component and the predicted TDOA pdf. Finally, (3) the enhanced TDOA is obtained. We evaluate the proposed approach on the AV16.3 corpus [7], a real corpus with different motion scenarios, using different tracking algorithms.

This work was partly supported by the European Union through the Marie-Curie Initial Training Network (ITN) SCALE (Speech Communication with Adaptive LEarning, FP7 grant agreement number 213850); by the Federal Republic of Germany, through the Cluster of Excellence for Multimodal Computing and Interaction (MMCI); and by Swiss National Science Foundation through the Swiss National Center of Competence in Research (NCCR) on Interactive Multimodal Information Management (IM2).

Our studies show that the proposed approach significantly reduces the angular error when compared to conventional approaches.

The paper is organized as follows. Section 2 provides a brief overview on acoustic source tracking problem. Section 3 presents the proposed approach. Section 4 presents the experimental results. Finally, in Section 5 we conclude.

2 Acoustic Source Tracking Problem

The arrival of sound waves at an array of microphones introduces time differences between the individual sensor pairs. This happens in dependence of the angle of arrival – that is, the azimuth θ and elevation ϕ – as well as the positions m_i, $i = 1, \ldots, M$ of the microphones. Under the far field assumption, in which the distance of the source from the microphones is neglected, the TDOA at the (i, j)-th sensor pair (m_i, m_j) can be calculated as:

$$\tau_{i,j}\left(p[\theta,\phi]\right) = \frac{p[\theta,\phi]^T (m_j - m_i)}{c} \tag{1}$$

where c denotes the speed of sound and $p[\theta, \phi]$ denotes the direction of arrival $\left[\cos(\phi)\sin(\theta),\ \cos(\phi)\cos(\theta),\ \sin(\phi)\right]^T$. Source localization approaches may use these time differences by either

(a) constructing a spatial filter (beamformer), which scans all possible source locations, and then taking that position where the signal energy is maximized [5].

(b) using a two stage approach, which consists in first estimating the TDOAs of all considered microphone pairs and then inferring the most likely source position [2, 3].

As our approach falls into the second category we proceed by first reviewing the Bayesian tracking framework in Section 2.1, explaining TDOA estimation in Section 2.2. Section 2.3 finally elaborates on how source localization can be performed based on estimated TDOAs.

2.1 Bayesian Tracking Framework

The problem of tracking a time-varying system state x_t based on a sequence $y_{1:t} = \{y_1, \ldots, y_t\}$ of corresponding observations is usually formulated as a Bayesian estimation problem in which

- Step 1: a process model $x_t = f(x_{t-1}, v_t)$ is used to construct a prior $p(x_t|y_{1:t-1})$ for the state estimation problem at time t.
- Step 2: the joint predictive distribution $p(x_t, y_t|y_{1:t-1})$ of state and observation is constructed according to a measurement model $y_t = h(x_t, w_t)$ with *measurement noise* w_t.
- Step 3: the posterior distribution $p(x_t|y_{1:t})$ is obtained by conditioning the joint predictive density $p(x_t, y_t|y_{1:t-1})$ on the realized (actually measured) observation $Y_t = y_t$.

The first step is accomplished by transforming the joint random variable of the last state X_{t-1} and process noise V_t according to f: $X_t = f\left(X_{t-1}, V_t\right)$. In step 2, the joint distribution of X_t and Y_t is constructed by transforming (X_t, W_t) according to the augmented measurement function $\tilde{h}$ [8]:

$$\begin{bmatrix} X_t \\ Y_t \end{bmatrix} = \tilde{h}\left(\begin{bmatrix} X_t \\ W_t \end{bmatrix}\right) \quad \text{with} \quad \tilde{h}\left(\begin{bmatrix} x_t \\ w_t \end{bmatrix}\right) \triangleq \begin{bmatrix} x_t \\ h(x_t, w_t) \end{bmatrix}.$$

Recursion of the above mentioned transformations form the Bayesian tracking framework. The posterior filtering distribution $p(x_t|y_{1:t})$ constitutes the complete solution to the sequential probabilistic inference problem and allows us to calculate any optimal estimate of the state. Although this approach is simple, the optimal solution is usually tractable only for linear and Gaussian systems, in this case, all the involved random variables remain Gaussian at all times and the posterior can be obtained as a conditional Gaussian distribution [8]. This analytical closed form solution is generally known as the *Kalman filter* (KF). Most real-world systems, however, are nonlinear or/and non-Gaussian, therefore the optimal solution is intractable and approximate solutions must be used. These include well-known extensions of the Kalman filter, such as the *Unscented Kalman Filter* (UKF) [2], the *Extended Kalman Filter* (EKF) [3], *sequential Monte-Carlo methods* (particle filters) [4, 6] and *Gaussian sum filters* [9, 10].

2.2 GCC-Based TDOA Estimation

The most popular approach to estimate the TDOA between two microphones m_i and m_j is to use the generalized cross-correlation (GCC) with Phase Transform (PHAT) weighting [1]. This approach is based on calculating the correlation of the signals $s_i(t)$ and $s_j(t)$, which have been received at the microphones, according to:

$$\mathcal{R}_{i,j}(\tau) = \frac{1}{2\pi} \int_0^{2\pi} \frac{S_i(w) S_j^*(\omega)}{\left| S_i(\omega) S_j^*(w) \right|} e^{j\omega\tau} \mathrm{d}\omega \tag{2}$$

where $S_i(\omega)$ and $S_j(\omega)$ denote the short-time Fourier transforms of $s_i(t)$ and $s_j(t)$, respectively, and where $\mathcal{R}_{i,j}$ is their weighted cross correlation. Subsequently, the most "likely" TDOA $\widehat{\tau}_{i,j}$ is extracted as:

$$\widehat{\tau}_{i,j} = \mathrm{argmax}_\tau \, \mathcal{R}_{i,j}(\tau) \tag{3}$$

2.3 Acoustic Source Tracking Based on Estimated TDOAs

Once the TDOA has been estimated for a number of N microphone pairs, acoustic source tracking can be performed with any algorithm from section 2.1 (e.g., [2, 3, 9]). In order to do this, we use the following process model for tracking the azimuth θ and elevation ϕ of the source:

$$\begin{bmatrix} \theta_t \\ \phi_t \end{bmatrix} = f\left(\begin{bmatrix} \theta_{t-1} \\ \phi_{t-1} \end{bmatrix}, v_t \right) = \begin{bmatrix} \theta_{t-1} + v_{t,\theta} \\ \phi_{t-1} + v_{t,\phi} \end{bmatrix} \tag{4}$$

where $v_{t,\theta}$ and $v_{t,\phi}$ denote zero-mean Gaussian process noise with a variance of σ_θ^2 and σ_ϕ^2, respectively. Similar to the approaches taken in [2, 3, 6], we use

$$y_t = h\left(\begin{bmatrix} \theta_t \\ \phi_t \end{bmatrix}, \mathbf{w}_t \right) = \begin{bmatrix} \tau_{i_1,j_1}\left(p[\theta_t, \phi_t]\right) + w_{t,1} \\ \vdots \\ \tau_{i_N,j_N}\left(p[\theta_t, \phi_t]\right) + w_{t,N} \end{bmatrix} \tag{5}$$

as a measurement model. In this equation, $\tau_{i_n,j_n}\left(p[\theta_t, \phi_t]\right)$ denotes the predicted TDOA of the n-th microphone pair (i_n, j_n), with $n = 1, \ldots, N$, whereas $w_{t,n}$ is zero-mean Gaussian measurement noise with a variance of σ_W^2. This measurement model is nonlinear, as the calculation of the predicted TDOAs according to (1) involves evaluating sines and cosines for the direction of arrival $p[\theta_t, \phi_t]$. Hence, the tracking should be performed using one of the approximate solutions in section 2.1.

3 TDOA Gaussian Mixture Model

GCC-based TDOA estimation works well in an environment that is characterized by low noise and reverberation. However, as mentioned earlier in Section 1, it breaks down in moderately reverberant conditions where, early reflections and reverberation corrupt the GCC function through smearing and through introduction of secondary peaks.

Interpreting the normalized GCC as a probability distribution of the TDOA, similar as originally proposed in [6] and firstly applied in [4] for a steered response power (SRP) approach [5], allows a probabilistic approach to the problem of TDOA estimation. Given this interpretation, the maximal peak of the GCC can be considered to be the *maximum* estimate. This has been implicitly used in [2, 3]. In this work, we continue along these lines and propose a probabilistic approach, which tries to enhance each TDOA by a) approximating the TDOA pdf by a GMM, as described in Section 3.1, b) updating the GMM with knowledge that has been obtained in the tracking stage, as explained in Section 3.2, and c) finally, estimating the TDOA (Section 3.3).

Besides the multimodality of the GCC function, the choice of the GMM as approximation of the TDOA pdf is also motivated by the Gaussianity assumption of the tracking information, which makes its integration into the TDOA estimation stage easier and more reliable.

3.1 Gaussian Mixture Model

The most popular approach to estimate a maximum likelihood GMM from given data is the Expectation-Maximization (EM) algorithm. Using the EM algorithm to estimate a GMM for each microphone pair at each

time t, however, would be computationally expensive. Thus, we use a computationally less expensive method that provides comparable results to those obtained with the EM algorithm.

Let K_t^n be the number of GCC peaks of the n^{th} microphone pair at time t, and let $y_t^n = \{\hat{\tau}_1^n, ..., \hat{\tau}_{K_t^n}^n\}$ and $w_t^n = \{w_1^n, ..., w_{K_t^n}^n\} = \{GCC(\hat{\tau}_1^n), ..., GCC(\hat{\tau}_{K_t^n}^n)\}$ be the corresponding TDOAs and GCC values, respectively. For ease of notation, the time index t and the microphone pair index n are dropped in the rest of paper. Then, we construct the GMM as follows:

1. Determine the K peaks of the GCC.
2. Determine the K blocks $\{B_1, ..., B_K\}$ corresponding to the different peaks. By block we mean the peak interval, which starts at its left foot and ends at the right foot (e.g., see Figure 1).
3. Calculate the Gaussian pdf associated to each block.
4. Normalize the GMM weights $\{w_1, ..., w_K\}$ (GCC peaks).

The Gaussian pdf $\mathcal{N}(\tau; \mu_k, \sigma_k^2)$ corresponding to the k^{th} block B_k and its mixture weight $\hat{w}_k$ are given by :

$$\mu_k = \hat{\tau}_k \tag{6}$$

$$\sigma_k^2 = \frac{\sum_{i/\tau_i \in B_k} GCC(\tau_i)(\tau_i - \mu_k)^2}{\sum_{i/\tau_i \in B_k} GCC(\tau_i)} \tag{7}$$

$$\hat{w}_k = \frac{w_k}{\sum_{i=1}^{K} w_i} \tag{8}$$

The statistical properties of the GCC function ensure that one of the peaks corresponds to the true TDOA. This fact justifies the choice of the TDOA and the GCC values of the peaks to be the means and the weights of the GMM, respectively. The main problem, however, is to find the peak which corresponds to the direct path. This problem is treated in the next section.

3.2 Update of The Gaussian Mixture Model Using Tracking Information

Previous works, though not directly related to the acoustic source tracking problem, have shown that the use of prior information about the measurements can efficiently improve measurement detection [11] (e.g., the "gating" approach). Along this line, we present in this section an approach for updating the GMM through use of information that has been obtained in the tracking stage. This is achieved by (1) calculating the predicted pdf of the TDOA (Step 2 in Section 2.1) as it is expected by the tracking algorithm; and then, (2) calculate the similarity scores between the predicted pdf and each component in the GMM where the similarity score reflects the probability that the component generates the true TDOA observation; and finally, (3) updating the mixture weights of GMM based on the similarity scores.

Let g_p and g_k be the predicted pdf of the TDOA and the k^{th} component of the GMM, respectively. For calculating similarity scores, we propose two different similarity measures (SMs):

$$SM_1(g_p, g_k) = \frac{1}{1 + KLD(g_p||g_k)} \tag{9}$$

$$SM_2(g_p, g_k) = \int \sqrt{g_p(x) g_k(x)}\, \mathrm{d}x \tag{10}$$

Where, $KLD(g_p||g_k)$ is the *Kullbak-Leibler Divergence* between the two Gaussians. The second SM is the *Bhattacharyya Coefficient* (BC)[12]. These two SMs have closed form solutions for Gaussian distributions.

Having calculated the SM for each component of the GMM, we update the weights of the GM before estimating the TDOA. The new weight $\bar{w}_k$ of the k^{th} component is given by

$$\bar{w}_k = \frac{\hat{w}_k SM(g_p, g_k)}{\sum_{i=1}^{K} \hat{w}_i SM(g_p, g_i)}. \tag{11}$$

The update step smoothes out the unlikely components and enhances the ones which are "close" to the predicted TDOA. This step can be seen as a "correction" of the GMM (e.g., see Figure 1).

3.3 TDOA Estimation

After the GMM update, the TDOA estimate can be obtained in two different ways – the *maximum estimate* from (12) and the *mean estimate* from (13):

$$\tau_{max} = \arg\max_{\tau} \sum_{k=1}^{K} \bar{w}_k \mathcal{N}(\tau; \mu_k, \sigma_k^2) \tag{12}$$

$$\tau_{mean} = \sum_{k=1}^{K} \bar{w}_k \mu_k \tag{13}$$

In any case, the estimated TDOA can be used in an arbitrary single observation acoustic source tracking approach. To construct the observation vector, we first estimate the TDOA $\bar{\tau}_t^k$ for each microphone pair, $n = 1, \ldots, N$, and then combine these individual estimates to form a joint measurement y_t, with $y_t = [\bar{\tau}_t^1, ..., \bar{\tau}_t^N]$.

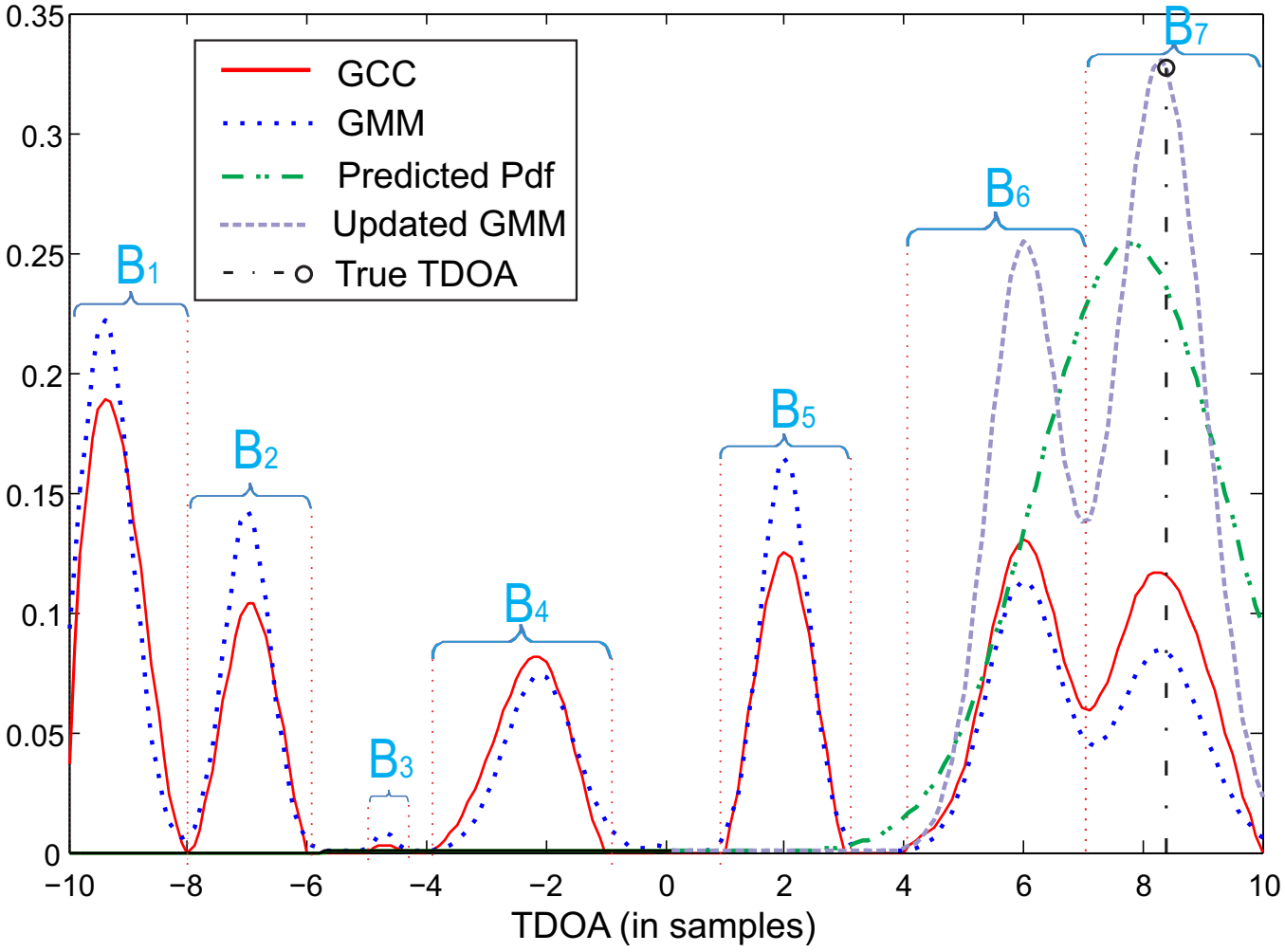

Fig. 1. *Illustration of the TDOA Gaussian mixture model and the Gaussian similarity measure (K = 7).*

Figure 1 illustrates the efficiency of the proposed method. The maximal GCC peak corresponds to a TDAO of -9 samples. The use of SM, however, alleviates the estimation error and recovers the true TDOA, which is 8.2 samples.

Sequence "seq11-1p-0100" / quickly moving				
tracking algorithm	root mean square error			
	azimuth	elevation	DOA	TDOA
UKF	5.56°	15.98°	16.92°	2.01
UKF+Gating	4.17°	7.12°	8.24°	1.05
UKF+SM	2.97°	4.92°	5.74°	0.64
SIR-PF	4.80°	10.33°	11.40°	2.01
SIR-PF+SM	3.29°	5.12°	6.09°	0.64
MH-PF	3.72°	5.94°	7.00°	—
MH-PF+SM	3.25°	4.81°	5.80°	—
MH-GMF	2.85°	4.25°	5.11°	—
MH-GMF+SM	3.21°	5.070°	5.99°	—
Sequence "seq02-1p-0000" / more stationary				
tracking algorithm	root mean square error			
	azimuth	elevation	DOA	TDOA
UKF	8.15°	20.23°	21.81°	2.33
UKF+Gating	2.71°	8.14°	8.58°	0.99
UKF+SM	2.83°	5.11°	5.84°	0.64
SIR-PF	7.54°	19.57°	20.98°	2.33
SIR-PF+SM	2.97°	5.46°	6.20°	0.62
MH-PF	3.99°	6.44°	7.58°	—
MH-PF+SM	3.32°	5.42°	6.36°	—
MH-GMF	2.71°	4.07°	4.89°	—
MH-GMF+SM	2.60°	3.86°	4.65°	—

Table 1. *Average root mean square error (RMSE) ,with and without Similarity Measure (SM), in azimuth, elevation and direction of arrival, with respect to the center of the array. The last column shows the average RMSE of the TDOA of 18 microphone pairs. This RMSE is calculated only for the single observation filters.*

4 Experiments and Results

4.1 Database and Experimental Setup

In order to evaluate the performance of the proposed algorithm, we performed a set of tracking experiments on the AV16.3 corpus [7]. In this corpus, real human speakers have been recorded in a smart meeting room (approximately 30m^2 in size) with a 20cm 8-channel circular microphone array. The sampling rate is 16 KHz and the real mouth position is known with an error of $\leq$ 5cm [7]. We present studies for two different sequences of this corpus: the highly non-stationary sequence "seq11-1p-0100", in which a single speaker is quickly moving in the room; and the relatively stationary sequence "seq02-1p-0000", in which a speaker is moving through 16 predefined locations while uttering one sentence "One,Two,Three,..." at each of the positions. These sequences are 32 and 185 seconds in length, respectively. The average distance of the speaker from the array is 1.18 and 1.53 meters, with a minimum of 0.57 and a maximum of 2.40 (links to the videos can be found in [7]).

The signal is divided into frames of 1024 samples (64ms). All the GCCs were calculated under use of PHAT [1] weighting. As there is no point in tracking an inactive speaker, we use a voice activity detector [13] for suppressing observations during silence frames. As a further precaution, the SM is replaced by gating [11] in the first frames and is used only after a duration T that ensures the true source is tracked.

In order to test the performance, we have combined the proposed method with 4 different algorithms that have been proposed as a solution to the acoustic source tracking problem: (i) the UKF [2] as well as a combination of the UKF with *Gating* [11], (ii) the Sequential Importance Resampling Particle Filter (SIR-PF) [4], which is implemented here as a TDOA-based approach, (iii) the Multiple Hypothesis Auxiliary Particle Filter (MH-PF) approach from [6], and (iv) the Multiple Hypothesis Gaussian Mixture Filter (MH-GMF) from [9]. UKF and PF are single observation

acoustic source tracking approaches which use the TDOA estimates from (12). In case of the MH-PF and MH-GMF the GCC is replaced by the updated GMM. The results are presented with and without the proposed approach using the second SM, i.e. BC. The use of KLD gives similar results.

4.2 Results and Analysis

Table 1 shows clearly that the integration of the prior information of the TDOA into the detection stage, through Gating or through the proposed approach denoted as "SM", improves the TDOA estimation and thereby, the tracking performance. The results also show that the proposed approach improves the performance of almost all the tracking algorithms, except the MH-GMF when it is applied to the sequence "seq11-1p-0100". This exception is due to the measurement model Eqn. (5), which assumes the source is stationary, whereas the speaker in this sequence is quickly moving. We can also conclude that the use of this approach is more relevant with single observation tracking algorithms, where the DOA error is 66% and 73% lower for the UKF and 46% and 70% lower for the PF. This compares to 17% and 16% improvement for the MH-PF and only 4% for the MH-GMF when it is applied to sequence "seq02-1p-0000". The difference in the improvement was expected, regarding that the multiple hypothesis filters propose to overcome the multimodality problem by considering multiple peaks with equal weights, whereas the SM assigns a likelihood weight to each Gaussian before estimating the observations, and thereby, improves the TDOA estimates. We can also notice that, with SM, the performance of the single observation filters, which are computationally more efficient, is close to the performance of the multiple observations filters. This makes the former more attractive.

Table 1 also shows that the reason behind the improvement is the reduction of the TDOA root means square error, which is around 0.63. This value is compared to the inherent 0.5 samples precision error due to the GCC method. Although this could be slightly improved through GCC interpolation, the gain we obtained from this was negligible.

5 Conclusions

We presented a Gaussian mixture model of the TDOA which couples detection stage and tracking stage to enhance TDOA estimates. More specifically, our study shows that the proposed model can be efficiently used to improve the performance of acoustic source tracking algorithms, as it reduces the problem of erroneous TDOA estimates by using the prior information given by the predicted pdf of the TDOA. In this work, our focus was on single source tracking problem. Future work will investigate the generalization of this approach to multiple source tracking problem.

6 References

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

[2] S. Gannot and T. G. Dvorkind, "Microphone array speaker localizers using spatial-temporal inforamtion," *EURASIP Journal on Applied Signal Processing*, 2006.

[3] U. Klee, T. Gehrig, and J. McDonough, "Kalman filters for time delay of arrival-based source localization," *EURASIP Journal on Applied Signal Processing*, 2006.

[4] D. B. Ward and R. C. Williamson, "Particle filter beamforming for acoustic source localization in a reverberant environment," in *Proc. ICASSP*, 2002.

[5] J. H. DiBiase, *A high-accuracy, low-latency technique for talker localization in reverberant environments using microphone arrays*, Ph.D. thesis, Brown University, 2000.

[6] J. Vermaak and A. Blake, "Nonlinear filtering for speaker tracking in noisy and reverberant environments," in *Proc. ICASSP*, 2001.

[7] G. Lathoud, J.-M. Odobez, and D. Gatica-Perez, "AV16.3: An audio-visual corpus for speaker localization and tracking," in *Proc. MLMI 04 Workshop*, 2006, pp. 182–195.

[8] F. Faubel and D. Klakow, "A transformation-based derivation of the Kalman filter and an extensive unscented transform," in *Proc. IEEE Workshop on SSP*, 2009, pp. 161–164.

[9] Y. Oualil, F. Faubel, and D. Klakow, "A multiple hypothesis Gaussian mixture filter for acoustic source localization and tracking," Idiap-RR Idiap-RR-09-2012, Idiap, 2012, Submitted to IEEE SSP Workshop 2012.

[10] D. L. Alspach and H. W. Sorenson, "Nonlinear bayesian estimation using gaussian sum approximations," *IEEE Transactions on Automatic Control*, vol. 17, no. 4, pp. 439–448, 1972.

[11] Y. Bar-Shalom and T. E. Fortmann, *Tracking and Data Association*, Academic Press, 1988.

[12] T. Kailath, "The divergence and Bhattacharyya distance measures in signal selection," *IEEE Transactions on Communication Theory*, vol. 15, pp. 52–60, 1967.

[13] J. Sohn, N. S. Kim, and W. Sung, "A statistical model-based voice activity detection," *IEEE Signal Process. Lett.*, vol. 6, no. 1, pp. 1–3, 1999.