CCIT Report # 469 February 2004

On the Use of Spatial-Temporal Information in Speaker Localization Applications

Tsvi Gregory Dvorkind and Sharon Gannot

Abstract

Speaker localization based on a microphone array and using a dual step approach is addressed. The first stage, which is not the main concern of this paper, is comprised of estimating the *time difference of arrival* between the speech signal received by each microphones pairs. These readings are then used by the second stage for the actual localization. The speaker's smooth trajectory is used for improving the current position estimate. Two localization schemes, which exploit the temporal information, are presented. The first is the *Extended Kalman filter*. The second is a recursive form of the Gauss method. Experimental study as well as approximate analytical evaluation supports the potential of the proposed methods.

Keywords

Talker Localizer; Extended Kalman filter; Gauss method; Cramér-Rao Lower Bound

I. Introduction and Problem Formulation

Determining the spatial position of a speaker finds a growing interest in video conference scenario where automated camera steering and tracking are required. In this work we address approaches for determining speaker position which are comprised of two stages. In the first stage, the TDOA is estimated using spatially separated microphone pairs (e.g. [1],[2], [3], and [4]). In the second stage, these readings are used for the actual localization (e.g. [5], [6] and [7]). These methods exploit the spatial information obtained by different microphone pairs, but do not exploit the temporal information available from adjoint speaker position estimates. This information is relevant for the current position estimate, due to the speaker smooth trajectory. A shorter preliminary conference version of these ideas has been published in [8].

Consider an array of $M+1$ microphones, placed at the Cartesian coordinates $\underline{m}_i \triangleq [x_i, y_i, z_i]^T$; $i = 0, \ldots, M$ where $\underline{m}_0 = [0,0,0]^T$ is the reference microphone, placed at the axes origin and $(\cdot)^T$ stands for the transpose operation. Define the source coordinate at time instance t by $\underline{s}(t) \triangleq [x_s(t), y_s(t), z_s(t)]^T$. Each of the M microphones, combined with the reference microphone, is used at time instance t to produce a TDOA measurement $\tau_i(t)$; $i = 1 \ldots M$. Denote the i-th **range difference** measurement by $r_i(t) = c\tau_i(t)$, where c

T. G. Dvorkind is with the Department of Electrical Engineering, Technion - Israel Institute of Technology, Technion City, Haifa 32000, Israel (email: dvorkind@tx.technion.ac.il).

S. Gannot is with the School of Engineering, Bar-Ilan University, Ramat-Gan 52900, Israel (email: gannot@eng.biu.ac.il)
Corresponding author: S. Gannot

is the sound propagation speed (approximately 340[m/s] in air). The non-linear equations for estimating the source location parameters $\underline{s}(t)$ are (see for example [7]):

$$\mathbf{A}(t)\underline{f}(\underline{s}(t)) \approx \underline{d}(t) \tag{1}$$

where $\underline{f}^T(\underline{s}(t)) \triangleq \left[\underline{s}^T(t), \|\underline{s}(t)\|\right]$ and

$$\mathbf{A}(t) \triangleq \begin{bmatrix} \underline{m}_1^T, r_1(t) \\ \vdots \\ \underline{m}_M^T, r_M(t) \end{bmatrix}, \quad \underline{d}(t) \triangleq \frac{1}{2} \begin{bmatrix} \|\underline{m}_1\|^2 - r_1^2(t) \\ \vdots \\ \|\underline{m}_M\|^2 - r_M^2(t) \end{bmatrix}.$$

Note, that only approximate equality holds in (1), since the range difference measurements are noisy. Our goal is to estimate the source position $\underline{s}(t)$ from the noisy measurements.

The organization of the rest of the paper is as follows. In Section II we derive Gauss and recursive Gauss (RG) solutions for the localization problem. A Baysian approach, namely the extended Kalman filter (EKF), is presented in Section III. The equivalence of RG and EKF approaches is discussed in Section IV. Test cases are presented in Section V.

II. Gauss and Recursive Gauss Algorithms

Huang *et al.* [7] addressed the non-linear set in (1) and solved it by using Lagrange multiplier. Since a polynomial of degree six is involved in the proposed method (denoted *linear-correction least-squares* (LCLS)), no closed-form solution exists. Thus, the iterative secant method is used for the root search.

A. Gauss Solution

The solution of (1) presented by [7] involves iterations. We suggest to mitigate the non-linearity by an alternative method, i.e the Gauss method. Note, that Eq. (1) becomes a (non-linear) *Least Squares* (LS) problem if the number of microphone pairs fulfills $M > 3$, i.e there are more equations than unknowns. The resulting non-linear LS problem can be solved by applying the Gauss method. Define $\underline{s}^{(l)}(t)$, the estimate of $\underline{s}(t)$ at the l-th iteration. Define also $\underline{h}_t(\underline{s}^{(l)}(t)) \triangleq \mathbf{A}(t)\underline{f}(\underline{s}^{(l)}(t))$, and the associated gradient matrix by $\mathbf{H}_t(\underline{s}^{(l)}(t)) = \nabla_{\underline{s}}\underline{h}_t(\underline{s}^{(l)}(t))$. By applying first-order approximation to $\underline{h}_t(\underline{s}(t))$, the Gauss iterations take the well known form

$$\underline{s}^{(l+1)}(t) = \underline{s}^{(l)}(t) + \left(\mathbf{H}_t(\underline{s}^{(l)}(t))^T \mathbf{H}_t(\underline{s}^{(l)}(t))\right)^{-1} \mathbf{H}_t(\underline{s}^{(l)}(t))^T \left(\underline{d}(t) - \underline{h}_t(\underline{s}^{(l)}(t))\right).$$

This solution exploits only the spatial information obtained by the separated microphone pairs at a specific time instance, but does not consider the temporal information.

B. Recursive Gauss (RG) Procedure

To exploit the temporal information, a recursive solution the Gauss method must be obtained. We begin by evaluating Eq. (1) at all measurements from $t = 1$ till $t = N$:

$$\mathbf{A}(t=1)\underline{f}(\underline{s}(1)) \approx \underline{d}(1), \; \ldots, \; \mathbf{A}(t=N)\underline{f}(\underline{s}(N)) \approx \underline{d}(N). \tag{2}$$

Note that this is still a nonlinear equation set in the unknown positions $\underline{s}(t)$; $t = 1, \ldots, N$, due to the nonlinearity introduced by $\underline{f}$. By assuming that $\underline{s}(t)$ is slowly varying with time, a recursive solution can be derived.

The proposed method starts by resolving the nonlinearities using first-order approximation (as with the original Gauss method), and then continues by deriving a recursion (applying further approximation). This solution will be referred to as *Recursive Gauss* (RG).

Consider a nonlinear equation set for the unknown $p \times 1$ parameter vector $\underline{\theta} \in \mathcal{C}^p$:

$$\underline{h}_{1:N}(\underline{\theta}) = \underline{d}_{1:N}$$

where $\underline{h}_{1:N}^T(\underline{\theta}) \triangleq \left[\underline{h}_1^T(\underline{\theta}) \cdots \underline{h}_N^T(\underline{\theta})\right]$ and $\underline{d}_{1:N}^T \triangleq \left[\underline{d}_1^T \cdots \underline{d}_N^T\right]$. $\underline{h}_t$ and $\underline{d}_t$ are K nonlinear equations and K measurements, available at time instance t, respectively. Applying first-order approximation around an initial guess $\underline{\theta}^{(0)}$ (as with the Gauss method) we obtain:

$$\underline{h}_{1:N}(\underline{\theta}^{(0)}) + \mathbf{H}_{1:N}(\underline{\theta}^{(0)})\left(\underline{\theta} - \underline{\theta}^{(0)}\right) \approx \underline{d}_{1:N} \tag{3}$$

where $\mathbf{H}_{1:N}$ is the $NK \times p$ gradient matrix:

$$\mathbf{H}_{1:N}^T(\underline{\theta}) \triangleq \left[\mathbf{H}_1^T(\underline{\theta}) \cdots \mathbf{H}_N^T(\underline{\theta})\right].$$

$\mathbf{H}_t(\underline{\theta}) = \nabla_{\underline{\theta}} \underline{h}_t(\underline{\theta})$ is the $K \times p$ gradient matrix of $\underline{h}_t(\underline{\theta})$. According to the Gauss method, the iterative LS solution to the linearized set (3) is:

$$\underline{\theta}^{(l+1)} = \arg\min_{\underline{\theta}} \left\| \underline{d}_{1:N} - \left(\underline{h}_{1:N}(\underline{\theta}^{(l)}) + \mathbf{H}_{1:N}(\underline{\theta}^{(l)})\left(\underline{\theta} - \underline{\theta}^{(l)}\right) \right) \right\|$$

where the superscript denotes the iteration number. Consider the next measurements $\underline{h}_{N+1}(\underline{\theta}) = \underline{d}_{N+1}$ available at time instance $N+1$. In order to estimate $\underline{\theta}$ we will use all the available measurements simultaneously. Though we could evaluate all $(N+1)K$ equations at the current estimate $\underline{\theta}^{(l+1)}$, we will do so **only** for the new equations. Namely, instead of minimizing in the LS sense the following residual norm

$$\min_{\underline{\theta}} \left\| \underline{d}_{1:N+1} - \left(\underline{h}_{1:N+1}(\underline{\theta}^{(l+1)}) + \mathbf{H}_{1:N+1}(\underline{\theta}^{(l+1)})\left(\underline{\theta} - \underline{\theta}^{(l+1)}\right) \right) \right\|$$

we will minimize a modified LS problem

$$\min_{\underline{\theta}} \left\| \begin{array}{ccc} \underline{d}_{1:N} & - & \left(\underline{h}_{1:N}(\underline{\theta}^{(l)}) + \mathbf{H}_{1:N}(\underline{\theta}^{(l)})\left(\underline{\theta} - \underline{\theta}^{(l)}\right) \right) \\ \underline{d}_{N+1} & - & \left(\underline{h}_{N+1}(\underline{\theta}^{(l+1)}) + \mathbf{H}_{N+1}(\underline{\theta}^{(l+1)})\left(\underline{\theta} - \underline{\theta}^{(l+1)}\right) \right) \end{array} \right\|.$$

The reason for this approximation is to keep past solutions intact, thus enabling a recursive solution to be derived. Now, using *stochastic approximation*, i.e. replacing the iteration index by the time index, a sequential algorithm is obtained. To summarize the procedure, an estimate for $\underline{\theta}$ at the current time instance t (denoted by $\hat{\underline{\theta}}(t)$) is obtained by solving the following LS problem sequentially using the *recursive LS* (RLS) procedure:

$$\hat{\underline{\theta}}(t) = \arg\min_{\underline{\theta}} \left\| \begin{bmatrix} \mathbf{H}_1(\hat{\underline{\theta}}(0)) \\ \vdots \\ \mathbf{H}_t(\hat{\underline{\theta}}(t-1)) \end{bmatrix} \underline{\theta} - \underline{y}_{1:t} \right\| \tag{4}$$

where

$$\underline{y}_{1:t} = \begin{bmatrix} \underline{y}_1 \\ \vdots \\ \underline{y}_t \end{bmatrix} \triangleq \begin{bmatrix} \underline{d}_1 - \underline{h}_1(\hat{\underline{\theta}}(0)) + \mathbf{H}_1(\hat{\underline{\theta}}(0))\hat{\underline{\theta}}(0) \\ \vdots \\ \underline{d}_t - \underline{h}_t(\hat{\underline{\theta}}(t-1)) + \mathbf{H}_t(\hat{\underline{\theta}}(t-1))\hat{\underline{\theta}}(t-1) \end{bmatrix}$$

with $\hat{\underline{\theta}}(0)$ the initial estimate for the parameter set. We note that in many practical situations the parameter set $\underline{\theta}$ might slowly vary with time. In these cases a common practice is to apply the RLS algorithm with a diagonal weight matrix that uses a forgetting factor $0 < \alpha \leq 1$ to weight past equations.

Another practical issue concerns the computational burden. At each time instance new K equations become available, resulting a $K \times K$ matrix inversion at each RLS iteration. However, by properly varying the forgetting factor α the computational complexity can be further reduced. This procedure is described in Appendix A.

C. Recursive Gauss (RG) Application

Denote the parameter set by $\underline{\theta} = \underline{s}$. Calculating the gradient matrix of the left hand side of (1) and the vector $\underline{y}_t$ as defined in (4) we obtain:

$$\mathbf{H}_t(\underline{s}) = \begin{bmatrix} \underline{m}_1^T + \frac{r_1(t)}{\|\underline{s}\|}\underline{s}^T \\ \vdots \\ \underline{m}_M^T + \frac{r_M(t)}{\|\underline{s}\|}\underline{s}^T \end{bmatrix}, \; \underline{y}_t = \underline{d}_t = \frac{1}{2} \begin{bmatrix} \|\underline{m}_1\|^2 - r_1^2(t) \\ \vdots \\ \|\underline{m}_M\|^2 - r_M^2(t) \end{bmatrix}$$

where we have used the fact that for the problem at hand $\mathbf{H}_t(\underline{\theta})\underline{\theta} = \underline{h}_t(\underline{\theta})$. Then $\hat{\underline{s}}(t)$ is evaluated by solving (4) with RLS and a forgetting factor $\alpha < 1$.

III. Extended Kalman Filter (EKF)

The non-linear set in Eq. (1) can be also solved in the Bayesian framework. The optimal *minimum mean square error* (MMSE) solution becomes complicated in this non-linear case, and sub-optimal solutions are called upon. Such a solution is the *extended Kalman filter* (EKF). As the actual movement model is not known in advance, we use a *random walk* model instead

$$\begin{cases} \underline{s}(t+1) &= \underline{s}(t) + \underline{w}(t) \\ \underline{r}(t) &= \underline{h}(\underline{s}(t)) + \underline{v}(t) \end{cases} \tag{5}$$

where $\underline{w}(t)$ is the state driving noise and $\underline{v}(t)$ is the measurement noise. $\underline{h}$ represents the nonlinear range difference measurement equations, given by:

$$\underline{h}(\underline{s}) \triangleq \begin{bmatrix} \|\underline{m}_1 - \underline{s}\| - \|\underline{s}\| \\ \vdots \\ \|\underline{m}_M - \underline{s}\| - \|\underline{s}\| \end{bmatrix}. \tag{6}$$

We note that the same approach was taken in [9], but in a different context.

IV. Equivalence of RG and EKF

It is well known that the RLS algorithm can be viewed as a special case of the Kalman filter. We show now that the same equivalence holds for the recursive Gauss algorithm, derived in Section II-B, and the **extended** Kalman filter. Using a diagonal weight matrix and setting the forgetting factor to α, the RG algorithm coincide with the EKF formulation for the following state-space model,

$$\begin{cases} \underline{\theta}(t+1) &= \underline{\theta}(t) \\ \underline{d}_t &= \underline{h}_t(\underline{\theta}(t)) + \underline{v}(t) \end{cases}$$

The equivalence holds when $\mathbf{R}(t) \triangleq \mathrm{Cov}(\underline{\mathrm{v}}(\mathrm{t})) = \alpha \mathbf{I}$ (where $\mathbf{I}$ stands for the identity matrix) and with the initial condition $\mathbf{P}_{0|-1} \triangleq \mathrm{Cov}(\underline{\theta}(0)) = \frac{1}{\alpha}\mathbf{P}(0)$. Moreover, this formulation is exactly the same as the one relating the RLS algorithm and (linear) Kalman filter.

V. Experimental Study

In this section we perform simulative comparison of several localization methods. To gain some insight on the obtainable performance of a microphone array with a small inter-element spread relative to the source position, calculation of the CRLB for a specific scenario is performed. This calculation leads us to a conclusion that the meaningful information lies in the azimuth and elevation angles estimates. We proceed by assessing four localization methods. Two of them are non-temporal (LCLS and Gauss iterations), and the other two (RG and EKF) exploit the temporal information.

A. Test Scenario

A set of 6 microphone pairs is placed on a sphere of radius $0.3[m]$ around a reference microphone placed at the origin, $\underline{m}_0 = [0, 0, 0]^T$, at the following positions:

$$\begin{aligned} \underline{m}_1^T &= \begin{bmatrix}0.3, \ 0, \ 0\end{bmatrix}, \quad \underline{m}_2^T = \begin{bmatrix}-0.3, \ 0, \ 0\end{bmatrix}, \ \underline{m}_3^T = \begin{bmatrix}0, \ 0.3, \ 0\end{bmatrix} \\ \underline{m}_4^T &= \begin{bmatrix}0, \ -0.3, \ 0\end{bmatrix}, \ \underline{m}_5^T = \begin{bmatrix}0, \ 0, \ 0.3\end{bmatrix}, \quad \underline{m}_6^T = \begin{bmatrix}0, \ 0, \ -0.3\end{bmatrix}. \end{aligned}$$

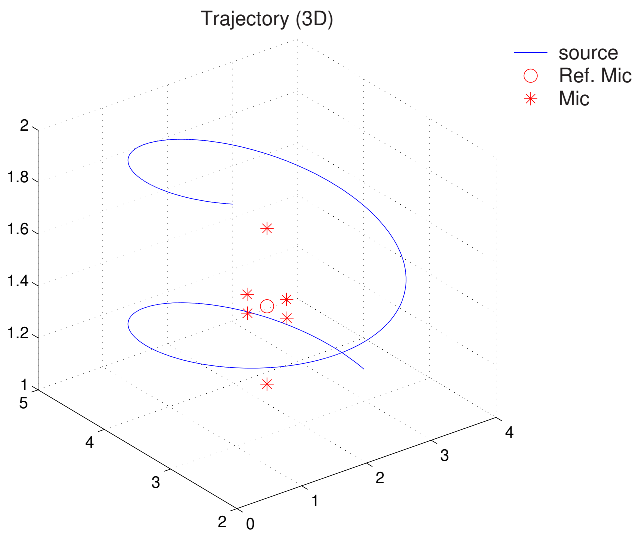

Fig. 1. Speaker trajectory

The speaker trajectory is set to be an helix with radius $R = 1.5$[m] around the reference microphone. The speaker movement speed is set to 0.5[m/s] and the total duration of the movement is $T = 30$[sec]. In Cartesian coordinates, the position of the speaker (in $\underline{m}_0$ coordinate system) as a function of time in the interval $t \in [0, T]$ is given by:

$$x(t) = R\cos(2\pi f t),\ y(t) = R\sin(2\pi f t),\ z(t) = \frac{t}{T} - 0.375$$

with $f = 0.0529[Hz]$. Along this trajectory, the overall change of the azimuth angle is within $\theta \in [0°, 570°]$ and of the elevation angle is within $\gamma \in [-14°, 22.5°]$. The entire scenario is depicted in Fig. 1.

B. The Cramér-Rao Lower Bound

We calculate now the CRLB for the tested scenario. We assume that the true range difference (or, equivalently, the TDOA) readings are contaminated by Gaussian distributed noise with zero-mean and *standard deviation* (STD) of $\sigma = 0.0425$[m]. This STD is equivalent to 1[sample] at a sample rate of $F_s = 8000$[Hz]. The existence of directional interferences and reverberation phenomenon might cause high level of noise correlation between microphone pairs and across time. Moreover, in high noise level the TDOA estimation algorithm might produce readings related to the directional noise source, causing multi-modal noise distribution. Nevertheless, for simplicity, we start by assuming (like Huang *et al.* [7]) that the noise is uni-modal (Gaussian) distributed spatially and temporally white. Under these conditions, the CRLB is calculated for both Cartesian and polar coordinates. The resulting bound (in meters, for the Cartesian coordinates and the radius, and in degrees for the azimuth and elevation angles) is depicted in Fig. 2. Note, that the Cartesian coordinates, as well as the radius, can not be accurately estimated in this scenario. This conclusion corresponds with the results presented in [7]. However the azimuth and elevation angles might be estimated in high accuracy. Fortunately, for camera steering applications, estimation of the azimuth and elevation angles suffices. Note

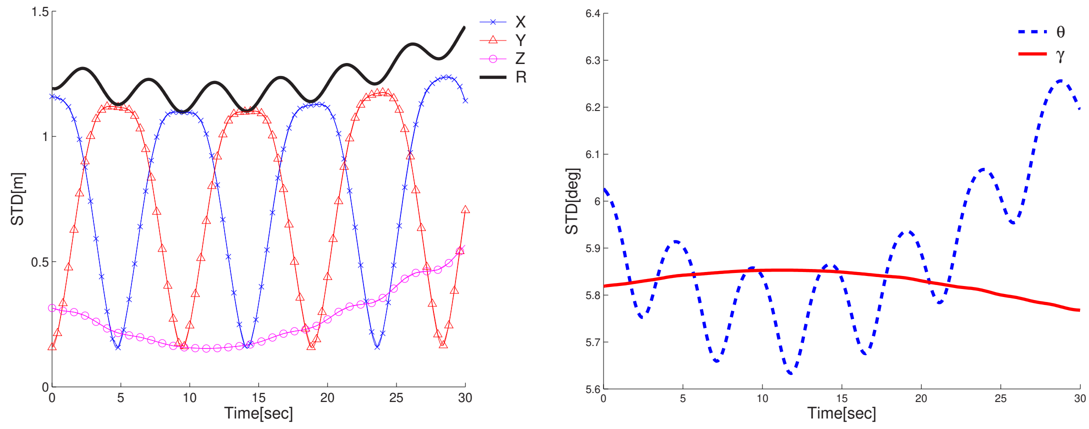

Fig. 2. CRLB results. Left: Cartesian coordinates and radius. Right: Azimuth (θ) and elevation (γ) angles.

also that the presented CRLB serves as a bound to the non-temporal methods alone, since past measurements are disregarded at each time instance.

C. Simulation Results

The previously presented setup is evaluated by four localization methods. The first is the *Linear Correction Least Squares* (LCLS) method, presented by Huang *et al.* [7]. The second is the Gauss method (denoted G) with 3 iterations at each time instance. The third is the recursive Gauss (RG) with forgetting factor $\alpha = 0.85$. The fourth is the EKF method evaluated with random-walk model and driving noise STD of 0.1[m] at each axis. The measurements covariance matrix is overestimated to $10\sigma^2\mathbf{I}$. 1000 Monte-Carlo trials are performed. The *Root Mean Square Error* (RMSE) of the angles estimate is presented in Fig. 3. As can be seen, the Gauss

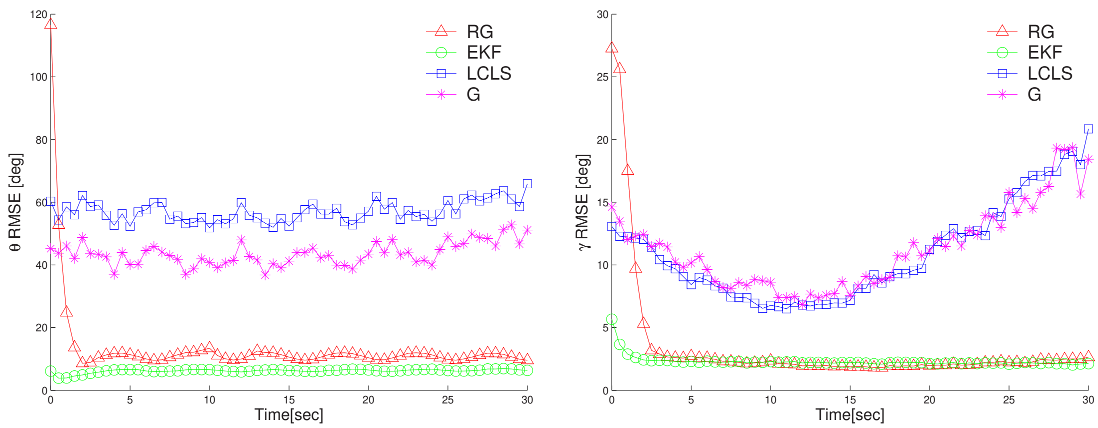

Fig. 3. RMSE results with white Gaussian noise. Left: RMSE for azimuth angle (θ). Right: RMSE for elevation angle (γ)

iterations and the LCLS method have comparable performance. However the RG and the EKF methods remarkably outperform them.

D. Switching Scenario

We proceed by testing a more realistic scenario. Consider the following simulation which is typical for a video conference scenario. Two speakers, located at two different and fixed locations alternately speak. The camera should be able to maneuver from one person to the other. For this scenario, using the same microphone positions as in the previous experiment, simulation is conducted with one speaker located at the polar position $[\theta = \frac{\pi}{4}[rad], \gamma = \frac{\pi}{4}[rad], R = 1.5[m]]$ and the other at $[\theta = \frac{3\pi}{4}[rad], \gamma = \frac{\pi}{3}[rad], R = 1.5[m]]$. A directional interference is placed at the position $[\theta = \frac{\pi}{2}[rad], \gamma = \frac{\pi}{4}[rad], R = 1[m]]$. Simulating reverberant conditions (reverberation time of $T_r = 0.25[sec]$) and mean SNR level of 10[dB]. Any method for TDOA extraction can be used in conjunction with our localization algorithm. However, to give specific simulations, we used TDOA readings, extracted from the noisy microphone data, by the RS1 algorithm described in [2], [3]. This method exploits the non-stationarity of the speech signal to estimate the ratio of the ATF-s relating he source signal and a pair of microphones. These noisy readings are then used by the localization methods to derive the angle estimate of the speakers. The same setup for the localization methods is applied here as well. Namely, the EKF localizer still uses the random walk model, though a better choice might have been asserted.

Figure 4 presents the azimuth angle estimates by the four methods. Figure 5 presents the elevation angle estimates. For this experiment, an angle estimate which diverts by more than $10°$ from its true trajectory is considered to be an anomaly and is not considered for the RMSE estimate. As can be seen from the plots, the temporal methods, especially the EKF algorithm, clearly outperforms the other methods. This is despite of the fact that the EKF is not using a valid state-space model. Note that the use of the random walk model in the EKF formulation explains the divergence from the RG method and is more appropriate for the tracking problem.

VI. Conclusions

We presented both non-temporal and temporal algorithms for talker localization and tracking. The Gauss method was shown to have comparable performance to the LCLS method. Two temporal methods were derived. One is within a non-Bayesian framework (RG algorithm) and the other is within the Baysian framework (EKF). The RG method is shown to be a degenerate case of the EKF. Evaluation of the CRLB showed that for a microphone array with a small inter-element spread relative to the source position, angle estimation might be obtained reliably (as opposed to the Cartesian coordinates estimates). Empirical results demonstrate the effectiveness of the use of the temporal information.

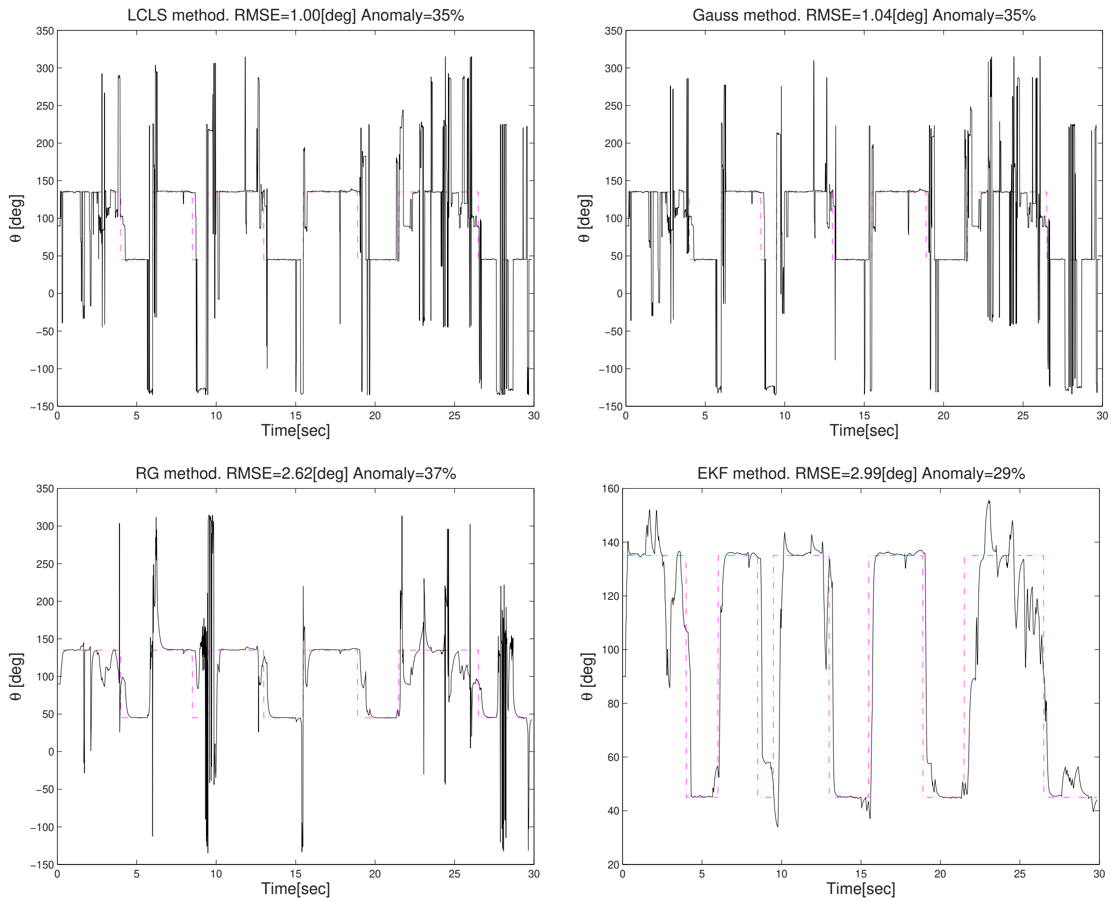

Fig. 4. Azimuth angle estimation results. Dotted line: True angle of the active speaker. Solid line: Estimation results. The method's name, anomaly percentage and RMSE of non-anomalous results are presented in the title of each plot.

Appendix

I. Recursive Least Squares for Multiple Readings

Assume a scenario in which for each time instance we have K scalar measurements $\underline{z}_t \in \mathbb{C}^K$ related to an unknown $p \times 1$ parameter vector $\underline{\theta} \in \mathbb{C}^p$ by a linear $K \times p$ transformation $\mathbf{H}_t$

$$\underline{z}_t \approx \mathbf{H}_t \underline{\theta}.$$

The approximation is due to the fact that the measurements are noisy, or due to slight modelling errors. N time instances can be augmented to a matrix form $\underline{z}_{1:N} \approx \mathbf{H}_{1:N}\underline{\theta}$ where

$$\underline{z}_{1:N} \triangleq \begin{bmatrix} \underline{z}_1 \\ \vdots \\ \underline{z}_N \end{bmatrix}; \; \mathbf{H}_{1:N} \triangleq \begin{bmatrix} \mathbf{H}_1 \\ \vdots \\ \mathbf{H}_N \end{bmatrix}.$$

The *weighted LS* (WLS) solution for $\underline{\theta}$, using nonnegative weight matrix $\mathbf{W}_{1:N}$ (of size $KN \times KN$) is:

$$\hat{\underline{\theta}} = \left(\mathbf{H}_{1:N}{}^{\dagger}\mathbf{W}_{1:N}\mathbf{H}_{1:N}\right)^{-1}\mathbf{H}_{1:N}{}^{\dagger}\mathbf{W}_{1:N}\underline{z}_{1:N} \tag{7}$$

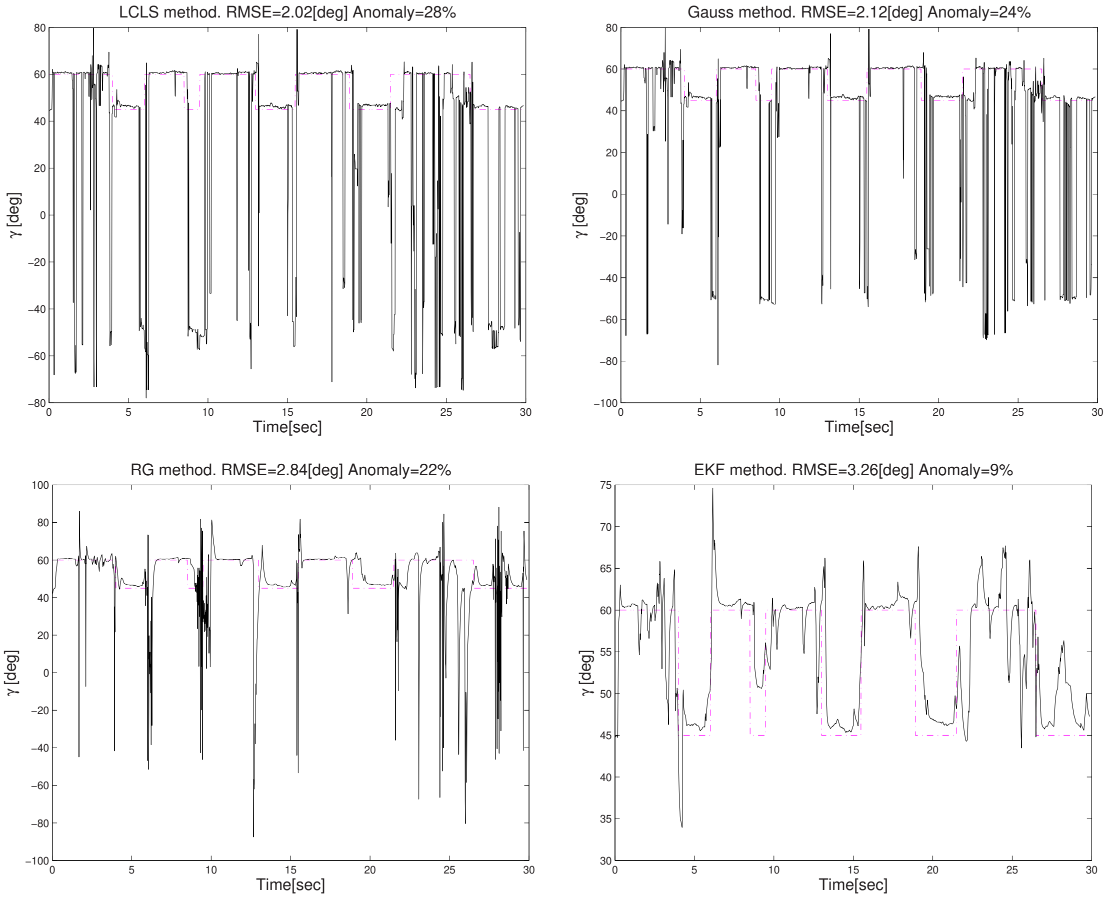

Fig. 5. Elevation angle estimation results. Dotted line: True angle of the active speaker. Solid line: Estimation results. The method's name, anomaly percentage and RMSE of non-anomalous results are presented in the title of each plot.

Our goal is to evaluate (7) recursively. If the parameters slowly change, a common approach is to apply a diagonal weight matrix $\mathbf{W}_{1:N}$ with powers of a forgetting factor $0 < \alpha \leq 1$ along its diagonal. Note, that for measurements associated with the same time instance, we wish to apply the same factor, since equations of the same time instance have equal importance. Such weight matrix can be represented recursively as:

$$\mathbf{W}_{1:N} = \begin{bmatrix} \alpha \mathbf{W}_{1:N-1} & \mathbf{0} \\ \mathbf{0}^{\dagger} & \mathbf{I} \end{bmatrix} ; \, \mathbf{W}_{1:1} = \mathbf{I}$$

where $\mathbf{I}$ and $\mathbf{0}$ stand for the identity and zero matrices of sizes $K \times K$ and $(N-1)K \times K$ respectively. Though it might seem that in order to derive a recursive solution for (7) a $K \times K$ matrix inversion should be made in each RLS iteration, in practice the complexity can be further reduced. This is obtained by applying the well known RLS algorithm with a minor twist. Consider a single equation which is updated into the recursion. We must check if this new equation belongs to the next time instance. If so, a memory factor $\alpha \leq 1$ is applied. If this is not the case and we are evaluating one of the K equations of the current time instance, a memory factor of 1 is used. Thus, in order to derive a recursion, where the update stage considers a **single** equation, the forgetting factor should vary. Notating the time instance by n and the sequential number of the equation by $nK + k$ (where $k \in \{1, \ldots, K\}$) the forgetting factor becomes

$$\text{forgetting factor} = \begin{cases} \alpha\ ; & k = 1, \\ 1\ ; & \text{otherwise.} \end{cases}$$

References

[1] T. Dvorkind and S. Gannot, "Speaker Localization in a Reverberant Environment," *IEEE proceedings, The 22nd convention of Electrical and Electronics Engineers in Israel.*, pp. 7–9, Dec. 2002.

[2] T. Dvorkind and S. Gannot, "Approaches for Time Difference of Arrival Estimation in a Noisy and Reverberant Environment," in *International Workshop on Acoustic Echo and Noise Control. Kyoto, Japan*, Sep 2003.

[3] T.G. Dvorkind and S. Gannot, "Time Difference of Arrival Estimation of Speech Source in Noisy and Reverberant Environment," CCIT report 457, Technion - Israel Institute of Technology, Dec. 2003.

[4] C.H. Knapp and G.C. Carter, "The Generalized Correlation Method for Estimation of Time Delay," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. 24, no. 4, pp. 320–327, 1976.

[5] J.O. Smith and J. S. Abel, "Closed-Form Least-Squares Source Location Estimation from Range-Difference Measurements," *IEEE transactions on Acoustics, Speech and Signal Processing*, vol. 35, no. 12, pp. 1661–1669, Dec. 1987.

[6] M. S. Brandstein, J. E. Adcock, and H. F. Silverman, "A Closed-Form Location Estimator for Use with Room Environment Microphone Arrays," *IEEE transactions on Speech and Audio Processing*, vol. 5, no. 1, pp. 45–50, Jan. 1997.

[7] Y. Huang, J. Benesty, G. W. Elko, and R. M. Mersereau, "Real-Time Passive Source Localization: A Practical Linear-Correction Least-Squares Approach," *IEEE transactions on Speech and Audio Processing*, vol. 9, no. 8, pp. 943–956, Jan. 2001.

[8] T. Dvorkind and S. Gannot, "Speaker Localization Exploiting Spatial-Temporal Information," in *International Workshop on Acoustic Echo and Noise Control. Kyoto, Japan*, Sep 2003.

[9] D. C. Popescu and C. Rose, "Emitter Localization in a Multipath Environment Using Extended Kalman Filter," in *33rd Conf. on Information Sciences and Systems - CISS*, The Johns Hopkins University, Baltimore, Maryland, USA, March 1999, vol. 1, pp. 147–150.