# Active Binaural Localization of Intermittent Moving Sources in the Presence of False Meaurements

Alban Portello[1], Patrick Danès[1] and Sylvain Argentieri[2]

***Abstract*— This paper takes place within the field of active sound source localization in a binaural context. A stochastic filtering strategy is presented for the localization of a still or moving source from a moving binaural sensor. The proposed method accounts for the source intermittence as well as for false measurements induced by the non-stationarity of the emitted signal. Its effectiveness is showed on experimental results.**

## I. Introduction

Robot audition can be defined as an artificial listening capability to recognize and understand the auditory environment. It covers various functionalities such as localization and separation of sound sources, and, at higher levels, speaker/speech recognition, multi-party interaction, etc. [1]. The first approaches were binaural, *i.e.* relied on two microphones mounted on a robotic head, and concerned low-level functions such as localization. The interest of using only two microphones is obvious in terms of cost and ease of implementation. Various techniques, relying on Interaural Intensity and Phase Differences (IID, IPD) computation together with Head Related Transfer Function (HRTF) exploitation, were put forward. A first idea was to model the effect of the head on the two perceived signals, leading to the Auditory Epipolar Geometry [2] or the Scattering Theory. But because of their limited performance and poor robustness w.r.t. modeling uncertainties and environment variability, binaural approaches were gradually supplanted by array processing [3]. Yet, it has recently been acknowledged that the limitations of a pair of microphones can be superseded by its mobility, which has given rise to the new topic of active binaural audition. Noticeably, emerging connections with recent theories of embodied cognition suggest alternative paradigms to perception in engineering [4].

Among the early contributions to active robot audition, the audio-visual tracker [5] integrates motor movements together with an adaptive ego-noise cancelling algorithm. More recently, the idea of fusing motion and perception appeared in [6][7]. This paper proposes an active binaural audition strategy to detect the activity of a speaker and localize him/her, in spite of outliers during cues extraction. It relies on the careful modeling developed in [8][9]. Its organization is as follows. First, the problem statement is recalled, together with the basic Kalman filter strategy [8]. Then, an extension coping with false measurements is exposed (Section III). This algorithm is adapted to intermittent sound sources in Section IV. Last, the approach is assessed on two scenarios.

*This work was conducted within the BINAAHR (BINaural Active Audition for Humanoid Robots) project funded by ANR (France) and JST (Japan) under Contract n°ANR-09-BLAN-0370-02.

[1]A. Portello and P. Danès are with CNRS, LAAS, 7 av. du colonel Roche, F-31400 Toulouse, France, and Univ. de Toulouse, UPS, LAAS; F-31400 Toulouse, France {aportell,danes}@laas.fr

[2]S. Argentieri is with UPMC Univ. Paris 06, UMR7222, ISIR, F-75005, Paris, France and CNRS, UMR7222, ISIR, F-75005, Paris, France argentieri@isir.upmc.fr

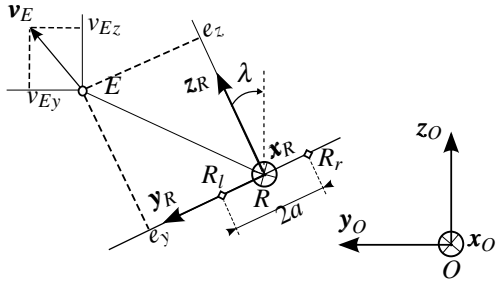



Fig. 1: The considered localization problem.

## II. Problem Statement and Basic Solution

### A. Problem statement

A pointwise sound emitter $E$ and a binaural sensor move independently on a common plane parallel to the ground. The two transducers equipping the sensor are denoted by $R_l$ and $R_r$. A frame $\mathscr{F}_R : (R, \boldsymbol{x}_R, \boldsymbol{y}_R, \boldsymbol{z}_R)$ is rigidly linked to the sensor, with $R$ the midpoint of the line segment $[R_l; R_r]$, $\boldsymbol{y}_R$ the vector $\frac{\boldsymbol{R}\boldsymbol{R}_l}{\|\boldsymbol{R}\boldsymbol{R}_l\|}$ and $\boldsymbol{x}_R$ the downward vertical vector. The frame $\mathscr{F}_E : (E, \boldsymbol{x}_O, \boldsymbol{y}_O, \boldsymbol{z}_O)$ attached to the source is parallel to the world reference frame $\mathscr{F}_O : (O, \boldsymbol{x}_O, \boldsymbol{y}_O, \boldsymbol{z}_O)$, with $\boldsymbol{x_O} = \boldsymbol{x_R}$ (see Fig. 1). $\|\boldsymbol{R_l}\boldsymbol{R_r}\| = 2a$ terms the transducers interspace. The source undergoes a translational motion (velocities $v_{Ey}, v_{Ez}$ of $\mathscr{F}_E$ w.r.t. $\mathscr{F}_O$ expressed along axes $\boldsymbol{y}_O, \boldsymbol{z}_O$), while the sensor is endowed with two translational and one rotational degrees-of-freedom (velocities $v_{Ry}, v_{Rz}$ of $\mathscr{F}_R$ w.r.t. $\mathscr{F}_O$ expressed along axes $\boldsymbol{y}_R, \boldsymbol{z}_R$; rotation velocity $\omega$ of $\mathscr{F}_R$ w.r.t. $\mathscr{F}_O$ around $\boldsymbol{x_O} = \boldsymbol{x_R}$). Assuming $v_{Ry}, v_{Rz}, \omega$ are known, the aim is to localize the emitter ($\mathscr{F}_E$) w.r.t. the binaural sensor ($\mathscr{F}_R$) from the sensed data at $R_l, R_r$. Free-field condition is assumed. The audio sensor is never localized w.r.t. $\mathscr{F}_O$.

### B. Mathematical modeling

*1) State space equation:* Without loss of generality, the relative attitude of $\mathscr{F}_R$ w.r.t. $\mathscr{F}_E$ can be described by a discrete-time stochastic state space equation of the form

$$\boldsymbol{X}_{[k+1]} = \boldsymbol{F}\boldsymbol{X}_{[k]} + \boldsymbol{G_1}\boldsymbol{u_1}_{[k]} + \boldsymbol{G_2}(\boldsymbol{X}_{[k]})\boldsymbol{u_2}_{[k]} + \boldsymbol{W}_{[k]} \qquad (1)$$

Therein, uppercase (resp. lowercase) vectors are random (resp. deterministic and known). $\boldsymbol{X} \triangleq (e_y, e_z, \lambda)'$ is the state vector to be estimated. It gathers the entries $e_y \triangleq \boldsymbol{R}\boldsymbol{E}.\boldsymbol{y_R}$ and

$e_z \triangleq \boldsymbol{RE}.\boldsymbol{z_R}$ of $\boldsymbol{RE}$ in $\mathscr{F}_R$, and the angle $\lambda \triangleq (\widehat{\boldsymbol{z_R}, \boldsymbol{z_O}})_{\boldsymbol{x_O}}$ (see Fig. 1). $\boldsymbol{W}_{[k]}$ is a Gaussian random dynamic noise with known statistics modeling uncertainty in the relative motion. The sensor velocities constituting $\boldsymbol{u_1} \triangleq (v_{Ry}, v_{Rz}, \omega)'$ are supposed known. Eq. (1) also assumes that the source velocities are available (*e.g.*, $\boldsymbol{u_2} \triangleq (v_{Ey}, v_{Ez})' = 0$ for a still source). If the source is moving at an unknown velocity, then $\boldsymbol{u_2}$ must be turned into a random variable $\boldsymbol{U_2}$ to be estimated, and Eq. (1) must be complemented to describe the dynamics of $\boldsymbol{U_2}$, *e.g.*, a random walk with known diffusion. More details about Eq. (1) are given in [8][9].

*2) Measurement equations:* In a conventional stochastic state space model, the measurement vector $\boldsymbol{z}$ is viewed as a sample of the measurement process $\boldsymbol{Z}$ linked to $\boldsymbol{X}, \boldsymbol{u_1}, \boldsymbol{u_2}$ by an output equation of the form

$$\boldsymbol{Z}_{[k]} = \boldsymbol{h}(\boldsymbol{X}_{[k]}, \boldsymbol{u_1}_{[k]}, \boldsymbol{u_2}_{[k]}) + \boldsymbol{V}_{[k]}, \tag{2}$$

with $\boldsymbol{V}$ a measurement noise. In our binaural approach, the Interaural Time Difference (ITD) [10] is used as the measurement. Importantly, when the source is moving, the equation relating the ITD to the source position is implicit, so that approximations are needed to get an explicit output equation like (2) [8][9]. When the source velocity is much lower than the sound speed, a "quasi-static" approximation of the ITD comes as

$$\text{ITD} = \frac{1}{c}\left(\sqrt{e_y^2+e_z^2+a^2+2ae_y} - \sqrt{e_y^2+e_z^2+a^2-2ae_y}\right). \tag{3}$$

The ITD is measured at each time from the raw audio data on the basis of Generalized Cross Correlation (GCC) techniques [11]. The GCC peak of the signals gathered at $R_l, R_r$ on a time window of length $T$ constitutes an estimator of the genuine ITD, whose properties have been studied in [12][13]. The associated Mean Square Error (MSE) depends on the Signal-To-Noise Ratio (SNR) and the Time-Bandwidth Product (TBP) of the binaural signals. When their values are sufficiently high, the measurement noise $\boldsymbol{V}$ can be assumed Gaussian zero-mean. In this case, the measurement is said "correct". Contrarily, when the SNR and/or the TBP fall below a given threshold, the extracted ITD is dominated by noise and does not bring any information about the source location. In this case, the measurement is said "false" and, assuming uncorrelated transducers noises, follows, say, a uniform probability density function (pdf) on $\mathscr{T} = [-T, T]$

$$\boldsymbol{z}_{[k]} \sim \mathscr{U}(\boldsymbol{Z}_{[k]}; \mathscr{T}). \tag{4}$$

In conclusion, at any time, the measurement is supposed to follow either (2)–(3) or (2)–(4). However, one must be aware that the real world exhibits measurements that are somewhere between false and correct. Also, the assumption of uncorrelated noises and source, which eases the problem of ITD estimation (as in [11][12][13]), is generally not satisfied in realistic environments (*e.g.* due to reverberation).

### C. Basic estimation strategy

Consider first that the source emits continuously a stationary signal such that at each time $k$ the measurement is correct. For example, say that the emitted signal is a sufficiently loud white noise. The nonlinear stochastic state space model (1)–(2) is thus considered. In this case, the square root Unscented Kalman Filter (sr-UKF) [14][15] provides sound and numerically robust approximations to the first two moments $\hat{x}_{[k|k]}, P_{[k|k]}$ and $\hat{x}_{[k|k-1]}, P_{[k|k-1]}$ of the pdfs $p(\boldsymbol{X}_{[k]}|\boldsymbol{Z}_{[1:k]} = \boldsymbol{z}_{[1:k]})$ and $p(\boldsymbol{X}_{[k]}|\boldsymbol{Z}_{[1:k-1]} = \boldsymbol{z}_{[1:k-1]})$ for a given sequence of measurements $\boldsymbol{z}_{[1:k]} = \boldsymbol{z}_{[1]}, \ldots, \boldsymbol{z}_{[k]}$. A major problem concerns the characterization of the initial state prior mean and covariance to be used in the sr-UKF initialization. When no knowledge about the initial state is available, a solution consists in initializing the sr-UKF with a flat prior (*e.g.* zero-mean and large covariance). However, in the considered problem, the propagation of widely spread distributions leads to overconfident conclusions. To avoid this problem, acknowledged in [8], an original multiple hypothesis filtering scheme (MH-srUKF) was proposed.

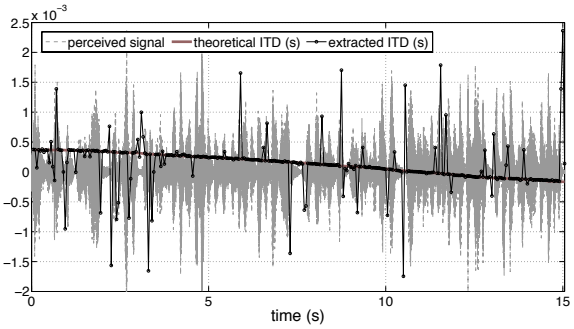



Fig. 2: Theoretical and extracted ITDs for speech signal.

### D. Pitfalls

During a speaker's utterance, due to the environment noise, the finite observation window, and the non-stationarity of speech, false measurements appear when the SNR/TBP are too low. To illustrate this point, Fig. 2 shows the theoretical and extracted ITDs along time for a moving sensor and a static loudspeaker emitting speech signal in an acoustically prepared room. ITDs were extracted every 50ms by GCC-PHAT [11], from 10ms-long audio records weighted by a Hanning window. The audio acquisition was performed at 44.1kHz, and the discrete GCC was interpolated using a lowpass filter to ensure good resolution. As depicted, most measurements fit well the theoretical ITD, in agreement with (2), while some of them corroborate (4). Obviously, the number of false measurements could be reduced by increasing the observation window size. However, this size must be chosen carefully, since the genuine ITD must not vary much within the observation window for the GCC estimation to be meaningful [11]. This is a fundamental concern when the source and sensor move. Consequently, the filtering strategy presented in [8], designed for an ideal source is not suited to speech signals. In fact, if the covariance of the measurement noise hypothesized in the MH-srUKF fitted its genuine value, then the false measurements, having a much larger diffusion, would compromise the filter stability. Of course, the MH-srUKF measurement noise covariance could be set to a larger value, so as to "capture" the false measurements

statistics, but such a solution would yield far too pessimistic conclusions. Finally, one could define a procedure that selects the measurements to be assimilated by the filter according to some criterion. For instance, the filter could be fed only with the measurements computed from binaural signals whose energy is greater than a predefined threshold. The problem is that with some probability, some validated measurements could in fact be false, so that the filter would assimilate them "believing" that they are correct [16]. So, a filter must be designed to handle false measurements in a probabilistic way. This is the topic of the next section.

## III. Dealing with False Measurements

### A. Probabilistic Data Association

A simplified version of the Probabilistic Data Association Filter (PDAF) [16] is derived. Define the random variable

$$I_{[k]} = \begin{cases} 1 & \text{if } \boldsymbol{z}_{[k]} \text{ is a correct measurement,} \\ 0 & \text{otherwise.} \end{cases} \quad (5)$$

Applying the total probability theorem w.r.t. the possible values of $I_{[k]}$, the posterior pdf of $\boldsymbol{X}$ at time $k$ writes as

$$\begin{aligned} p(\boldsymbol{X}_{[k]}|\boldsymbol{Z}_{[1:k]}=\boldsymbol{z}_{[1:k]}) &= \textstyle\sum_{i=0}^{1}\gamma_{i[k]}\, p(\boldsymbol{X}_{[k]}|I_{[k]}=i,\boldsymbol{Z}_{[1:k]}=\boldsymbol{z}_{[1:k]}) \\ &\approx \textstyle\sum_{i=0}^{1}\gamma_{i[k]}\,\mathcal{N}(\boldsymbol{X}_{[k]};\hat{x}_{i[k|k]},P_{i[k|k]}), \end{aligned} \quad (6)$$

where $\mathcal{N}(.;\hat{x},P)$ stands for the Gaussian pdf of mean $\hat{x}$ and covariance $P$, and $\gamma_{i[k]} \triangleq P(I_{[k]}=i|\boldsymbol{Z}_{[1:k]}=\boldsymbol{z}_{[1:k]})$ terms the posterior probability of each mode $i$. The posterior state pdf at time $k$ comes as a mixture of two Gaussian distributions: the posterior pdf assuming that $\boldsymbol{z}_{[k]}$ is correct, whose moments $\hat{x}_{1[k|k]}, P_{1[k|k]}$ can be computed from $\hat{x}_{[k-1|k-1]}, P_{[k-1|k-1]}$ through the UKF time and measurement updates, and the posterior pdf assuming that $\boldsymbol{z}_{[k]}$ is false, whose moments $\hat{x}_{0[k|k]}, P_{0[k|k]}$ can be computed from $\hat{x}_{[k-1|k-1]}, P_{[k-1|k-1]}$ by the UKF time update only. At each time $k$, each $i$-th mode likelihood $L_{i[k]} \triangleq p(\boldsymbol{Z}_{[k]}=\boldsymbol{z}_{[k]}|I_{[k]}=i,\boldsymbol{Z}_{[1:k-1]}=\boldsymbol{z}_{[1:k-1]})$ w.r.t. $\boldsymbol{z}_{[k]}$ come as

$$\begin{aligned} L_{1[k]} &= \mathcal{N}(\boldsymbol{z}_{[k]};\hat{z}_{1[k|k-1]},S_{1[k|k-1]}) \\ L_{0[k]} &= \mathscr{U}(\boldsymbol{z}_{[k]};\mathscr{T}), \end{aligned} \quad (7)$$

Therein, $\hat{z}_{1[k|k-1]}, S_{1[k|k-1]}$ are the moments of the predicted output pdf assuming $\boldsymbol{z}_{[k]}$ is correct. They can be deduced from the prior moments $\hat{x}_{[k|k-1]}, P_{[k|k-1]}$ in the UKF output prediction step by exploiting (2). The modes posterior probabilities are then deduced from the Bayes formula:

$$\gamma_{i[k]} = \frac{L_{i[k]}P(I_{[k]}=i|\boldsymbol{Z}_{[1:k-1]}=\boldsymbol{z}_{[1:k-1]})}{\sum_{l=0}^{1}L_{l[k]}P(I_{[k]}=l|\boldsymbol{Z}_{[1:k-1]}=\boldsymbol{z}_{[1:k-1]})}, i=0,1. \quad (8)$$

It is assumed that the probability to get a false or correct measurement at time $k$ is independent of the measurements at previous time steps, and is time-independent, so that $P(I_{[k]}=i|\boldsymbol{Z}_{[1:k-1]}=\boldsymbol{z}_{[1:k-1]})=P(I_{[k]}=i)=P_i$, with $P_0+P_1=1$. While the first assumption seems quite reasonable, the second one is less legitimate. Indeed, the false measurements rate is intuitively all the higher as the intensity of the binaural signals decreases. Nevertheless, as the measurement pdfs $p(\boldsymbol{Z}_{[k]}=\boldsymbol{z}_{[k]}|I_{[k]}=1,\boldsymbol{X}_{[k]}=\boldsymbol{x}_{[k]})$ and $p(\boldsymbol{Z}_{[k]}=\boldsymbol{z}_{[k]}|I_{[k]}=0,\boldsymbol{X}_{[k]}=\boldsymbol{x}_{[k]})$ have very distinct shapes (the first one shows a sharp mode, while the second ones spreads evenly over $\mathscr{T}=[-T,T]$) the assumed time-invariance of the false measurements probability does not influence much the results in practice. A mode-matched approximation of the posterior pdf then comes as

$$p(\boldsymbol{X}_{[k]}|\boldsymbol{Z}_{[1:k]}=\boldsymbol{z}_{[1:k]}) \approx \mathcal{N}(\boldsymbol{X}_{[k]};\hat{x}_{[k|k]},P_{[k|k]}), \quad (9)$$

$$\begin{aligned} \text{with } \hat{x}_{[k|k]} &= \textstyle\sum_{i=0}^{1}\gamma_{i[k]}\hat{x}_{i[k|k]} \quad (10) \\ P_{[k|k]} &= \textstyle\sum_{i=0}^{1}\gamma_{i[k]}\left(P_{i[k|k]}+(\hat{x}_{i[k|k]}-\hat{x}_{[k|k]})(\hat{x}_{i[k|k]}-\hat{x}_{[k|k]})^T\right). \end{aligned}$$

This scheme slightly differs from the original PDAF, in that only one measurement is available at each time, and no validation gate is introduced. Though, it will still be referred to as PDAF, along Algorithm 1.

### B. Multiple Hypothesis filter

While the PDAF can handle false measurements, running a single filter with a flat prior at initial time still leads to overconfident conclusions as acknowledged in [8]. To overcome this difficulty, the true initial state moments are supposed to belong to a finite set $\{\hat{x}^j_{[0|0]}, P^j_{[0|0]}\}_{j=1,\dots,J}$ defined from a partition of the admissible state space (*i.e.* of the admissible relative sensor-to-source locations) into $J$ overlapping cells $\{\mathscr{C}_j\}_{j=1,\dots,J}$, *e.g.* so that each 99 % probability ellipsoid defined from the Gaussian prior $\mathcal{N}(\boldsymbol{X}_0;\hat{x}^j_{[0|0]},P^j_{[0|0]})$ covers $\mathscr{C}_j$. Each hypothesis $F_j$ (*i.e.* the hypothesis that the $j^{\text{th}}$ initialization is correct) is assigned a given initial probability $W^j_{[0]} \triangleq P(F_j)$. So, the initial state prior pdf is described by the Gaussian mixture

$$p(\boldsymbol{X}_{[0]}) = \textstyle\sum_{j=1}^{J} W^j_{[0]}\mathcal{N}(\boldsymbol{X}_{[0]};\hat{x}^j_{[0|0]},P^j_{[0|0]}). \quad (11)$$

At time $k$, the posterior state pdf is obtained by applying the total probability theorem w.r.t. the measurement faithfulness and initialization hypotheses

$$\begin{aligned} &p(\boldsymbol{X}_{[k]}|\boldsymbol{Z}_{[1:k]}=\boldsymbol{z}_{[1:k]}) = \\ &\quad \textstyle\sum_{j=1}^{J} W^j_{[k]} \sum_{i=0}^{1}\gamma^j_{i[k]}\, p(\boldsymbol{X}_{[k]}|I_{[k]}=i,F_j,\boldsymbol{Z}_{[1:k]}=\boldsymbol{z}_{[1:k]}), \end{aligned} \quad (12)$$

where $W^j_{[k]} \triangleq P(F_j|\boldsymbol{Z}_{[1:k]}=\boldsymbol{z}_{[1:k]})$ is the posterior probability of $F_j$ and $\gamma^j_{i[k]} \triangleq P(I_{[k]}=i|F_j,\boldsymbol{Z}_{[1:k]}=\boldsymbol{z}_{[1:k]})$ is the posterior probability of $I_{[k]}=i$ assuming $F_j$. Eq. (12) can be approximated as

$$p(\boldsymbol{X}_{[k]}|\boldsymbol{Z}_{[1:k]}=\boldsymbol{z}_{[1:k]}) \approx \textstyle\sum_{j=1}^{J} W^j_{[k]}\mathcal{N}(\boldsymbol{X}_{[k]};\hat{x}^j_{[k|k]},P^j_{[k|k]}), \quad (13)$$

where the moments $\hat{x}^j_{[k|k]}, P^j_{[k|k]}$ are computed from a PDAF matched to $F_j$. So, a multiple hypothesis strategy handling false measurements can be derived by modifying the MH-srUKF proposed in [8] in two respects: the sr-UKFs matched to $\{F_j\}_{j=1,\dots,J}$ are replaced by PDAFs, and the likelihood $L^j_{[k]} \triangleq p(\boldsymbol{Z}_{[k]}=\boldsymbol{z}_{[k]}|F_j,\boldsymbol{Z}_{[1:k-1]}=\boldsymbol{z}_{[1:k-1]})$ of the filter initialized along $F_j$ is described by a mixture of Gaussian and uniform pdfs. This Multiple Hypothesis PDAF, hereafter referred to as MH-PDAF, is summarized in Algorithm 2.

**Algorithm 1** The PDAF.

$\left[\hat{x}_{[k|k]}, P_{[k|k]}, L_{[k]}\right] = \mathsf{PDAF}(\boldsymbol{z}_{[k]}, \hat{x}_{[k-1|k-1]}, P_{[k-1|k-1]}, \boldsymbol{u}_{[k-1]})$

1: **IF** $k=0$ **THEN** Define the initial conditions $\{\hat{x}_{[0|0]}, P_{[0|0]}\}$. **END IF**
2: **IF** $k \geq 1$ **THEN**
3: (UKF time update) Predict the moments $\hat{x}_{[k|k-1]}, P_{[k|k-1]}$ of $p(\boldsymbol{X}_{[k]}|\boldsymbol{Z}_{[1:k-1]}=\boldsymbol{z}_{[1:k-1]})$ from $\mathcal{N}(\boldsymbol{X}_{[k-1]}; \hat{x}_{[k-1|k-1]}, P_{[k-1|k-1]})$. The moments $\hat{x}_{0[k|k]}, P_{0[k|k]}$ of $p(\boldsymbol{X}_{[k]}|I_{[k]}=0, \boldsymbol{Z}_{[1:k]}=\boldsymbol{z}_{[1:k]})$ come as $\hat{x}_{0[k|k]}=\hat{x}_{[k|k-1]}$, $P_{0[k|k]}=P_{[k|k-1]}$.
4: (UKF measurement update) Predict the moments $\hat{z}_{1[k|k-1]}, S_{1[k|k-1]}$ of $p(\boldsymbol{Z}_{[k]}|I_{[k]}=1, \boldsymbol{Z}_{[1:k-1]}=\boldsymbol{z}_{[1:k-1]})$ from $\mathcal{N}(\boldsymbol{X}_{[k]}; \hat{x}_{[k|k-1]}, P_{[k|k-1]})$ by exploiting the correct measurement equation (2)–(3), and fuse $\boldsymbol{z}_{[k]}$ with $\hat{x}_{[k|k-1]}, P_{[k|k-1]}$ so as to get the moments $\hat{x}_{1[k|k]}, P_{1[k|k]}$ of $p(\boldsymbol{X}_{[k]}|I_{[k]}=1, \boldsymbol{Z}_{[1:k]}=\boldsymbol{z}_{[1:k]})$.
5: (Modes posterior probabilities) Evaluate $L_{[k]} \triangleq p(\boldsymbol{Z}_{[k]}=\boldsymbol{z}_{[k]}|\boldsymbol{Z}_{[1:k-1]}=\boldsymbol{z}_{[1:k-1]}) = \mathcal{N}(\boldsymbol{z}_{[k]}; \hat{z}_{1[k|k-1]}, S_{1[k|k-1]})P_1 + \mathscr{U}(\boldsymbol{z}_{[k]}; \mathscr{T})(1-P_1)$ and compute $\gamma_{0[k]} \triangleq P(I_{[k]}=0|\boldsymbol{Z}_{[1:k]}=\boldsymbol{z}_{[1:k]}) = \frac{\mathscr{U}(\boldsymbol{z}_{[k]}; \mathscr{T})(1-P_1)}{L_{[k]}}$ and $\gamma_{1[k]} \triangleq P(I_{[k]}=1|\boldsymbol{Z}_{[1:k]}=\boldsymbol{z}_{[1:k]}) = \frac{\mathcal{N}(\boldsymbol{z}_{[k]}; \hat{z}_{1[k|k-1]}, S_{1[k|k-1]})P_1}{L_{[k]}}$.
6: (Mode-matched moments) Output the posterior mean $\hat{x}_{[k|k]} = \sum_{i=0}^{1} \gamma_{i[k]} \hat{x}_{i[k|k]}$ and covariance $P_{[k|k]} = \sum_{i=0}^{1} \gamma_{i[k]} \left[P_{i[k|k]} + (\hat{x}_{i[k|k]} - \hat{x}_{[k|k]})(\hat{x}_{i[k|k]} - \hat{x}_{[k|k]})^T\right]$.
7: **END IF**

**Algorithm 2** The MH-PDAF.

$\left[\hat{x}_{[k|k]}, P_{[k|k]}, \{W^j_{[k]}\}_{j=1,\ldots,J}, \{\hat{x}^j_{[k|k]}, P^j_{[k|k]}\}_{j=1,\ldots,J}, \Lambda_{[k]}\right] = \mathsf{MH\text{-}PDAF}(\boldsymbol{z}_{[k]}, \{W^j_{[k-1]}\}_{j=1,\ldots,J}, \{\hat{x}^j_{[k-1|k-1]}, P^j_{[k-1|k-1]}\}_{j=1,\ldots,J}, \boldsymbol{u}_{[k-1]})$

1: **IF** $k=0$ **THEN** Define the moments $\{\hat{x}^j_{[0|0]}, P^j_{[0|0]}\}_{j=1,\ldots,J}$ of the state at initial time and the initial weights $\{W^j_{[0]} \triangleq P(F_j)\}_{j=1,\ldots,J}$. **END IF**
2: **IF** $k \geq 1$ **THEN**
3: **FOR** $j=1,\ldots,J$, **DO** (PDAF matched to $F_j$) $\left[\hat{x}^j_{[k|k]}, P^j_{[k|k]}, L^j_{[k]}\right] = \mathsf{PDAF}(\boldsymbol{z}_{[k]}, \hat{x}^j_{[k-1|k-1]}, P^j_{[k-1|k-1]}, \boldsymbol{u}_{[k-1]})$. **END FOR**
4: Compute $\Lambda_{[k]} \triangleq p(\boldsymbol{Z}_{[k]}=\boldsymbol{z}_{[k]}|\boldsymbol{Z}_{[1:k-1]}=\boldsymbol{z}_{[1:k-1]}) = \sum_{l=1}^{J} L^l_{[k]} W^l_{[k-1]}$.
5: **FOR** $j=1,\ldots,J$, **DO** Update the filters weights $W^j_{[k]} \triangleq P(F_j|\boldsymbol{Z}_{[1:k]}=\boldsymbol{z}_{[1:k]}) = \frac{L^j_{[k]} W^j_{[k-1]}}{\Lambda_{[k]}}$. **END FOR**
6: If some $W^j_{[k]}$ are lesser than a given threshold $\alpha$, then suppress the corresponding filters $F_j$. Decrease $J$ and renormalize all the weights accordingly.
7: **END IF**
8: (Mode-matched moments) Output the posterior mean $\hat{x}_{[k|k]} = \sum_{j=1}^{J} W^j_{[k]} \hat{x}^j_{[k|k]}$ and covariance $P_{[k|k]} = \sum_{j=1}^{J} W^j_{[k]} \left[P^j_{[k|k]} + (\hat{x}^j_{[k|k]} - \hat{x}_{[k|k]})(\hat{x}^j_{[k|k]} - \hat{x}_{[k|k]})^T\right]$.

## IV. A GLRT Approach to Detect Intermittent Sound Sources

This section shows how the MH-PDAF can be complemented with a detector of source activity/mute based on the Generalized Likelihood Ratio Test (GLRT).

### A. Intermittent source model

The following model can cope with an uttering speaker:

$$\mathscr{M}_1 : \begin{cases} \boldsymbol{X}_{[k+1]} = \boldsymbol{F}\boldsymbol{X}_{[k]} + \boldsymbol{G_1}\boldsymbol{u_1}_{[k]} + \boldsymbol{G_2}(\boldsymbol{X}_{[k]})\boldsymbol{u_2}_{[k]} + \boldsymbol{W}_{[k]} \\ \boldsymbol{Z}_{[k]} = \boldsymbol{h}(\boldsymbol{X}_{[k]}, \boldsymbol{u_1}_{[k]}, \boldsymbol{u_2}_{[k]}) + \boldsymbol{V}_{[k]} \quad \text{with prob. } P_1 \\ \boldsymbol{z}_{[k]} \sim \mathscr{U}(\boldsymbol{Z}_{[k]}; \mathscr{T}) \qquad\qquad \text{with prob. } 1\text{–}P_1. \end{cases} \tag{14}$$

When the speaker is mute, false measurements occur systematically. So, the model $\mathscr{M}_2$ for a silent speaker is identical to $\mathscr{M}_1$, excepted that the probability to get a correct measurement is zero. Under $\mathscr{M}_2$, the posterior state pdf is just the prediction

$$p(\boldsymbol{X}_{[k]}|F_j, \boldsymbol{Z}_{[1:k]}=\boldsymbol{z}_{[1:k]}) = \mathcal{N}(\boldsymbol{X}_{[k]}; \hat{x}^j_{0[k|k]}, P^j_{0[k|k]}), \tag{15}$$

where the moments $\hat{x}^j_{0[k|k]}, P^j_{0[k|k]}$ are computed from $\mathcal{N}(\boldsymbol{X}_{[k]}; \hat{x}^j_{[k-1|k-1]}, P^j_{[k-1|k-1]})$ with a UKF time update. Similarly, the modes posterior probabilities follow $W^j_{[k]} = W^j_{[k-1]}$. So, the algorithm MH-Pred suited to a mute speaker reduces to replacing, in the MH-PDAF algorithm, PDAFs with UKF time updates, and to suppressing step 6.

### B. Detection of the source state transitions

Sporadic jumps between $\mathscr{M}_1$ and $\mathscr{M}_2$ can be detected by a GLRT-based adaptive filtering scheme, in the vein of [17].

*1) Detection of a switch $\mathscr{M}_1 \to \mathscr{M}_2$:* Define the $N$-element data window $\mathscr{W} = \{k-N+1, \ldots, k\}$ ending at the current time $k$. Suppose that the set of moments and weights characterizing the Gaussian mixture approximation to the posterior pdf $p(\boldsymbol{X}_{[k-N]}|\boldsymbol{Z}_{[1:k-N]}=\boldsymbol{z}_{[1:k-N]})$ is known, and assume that the mode $\mathscr{M}_1$ was in effect until time $p \triangleq k-N+1$. Given $\boldsymbol{z}_{[p:k]}$, the aim is to define a decision rule to test the two following mutually exclusive hypotheses:

$\mathrm{H}_1(\theta_1)$ : a switch $\mathscr{M}_1 \to \mathscr{M}_2$ occurred at $\theta_1 \in \mathscr{W}$,
$\mathrm{H}_0$ : the system has remained in $\mathscr{M}_1$.

Note that $\mathrm{H}_1(\theta_1)$ is composite as the candidate switching time $\theta_1$ ranges over $\mathscr{W}$. Given the likelihoods $\Lambda_{\mathrm{H}_1}(\theta_1) \triangleq p(\boldsymbol{Z}_{[p:k]}=\boldsymbol{z}_{[p:k]}|H_1(\theta_1))$ and $\Lambda_{\mathrm{H}_0} \triangleq p(\boldsymbol{Z}_{[p:k]}=\boldsymbol{z}_{[p:k]}|H_0)$, the most likely hypothesis is detected according to the GLRT

$$\frac{\Lambda_{\mathrm{H}_1}(\hat{\theta}_1)}{\Lambda_{\mathrm{H}_0}} \underset{\mathrm{H}_0}{\overset{H_1}{\gtrless}} \eta, \tag{16}$$

where $\hat{\theta}_1 \triangleq \mathrm{argmax}_{\theta_1 \in \mathscr{W}} \left[\Lambda_{\mathrm{H}_1}(\theta_1)\right]$. The threshold $\eta$ is tuned so as to comply, say, with a selected false alarm probability. The likelihood of $\mathrm{H}_0$ can be expanded backwards as

$$\Lambda_{\mathrm{H}_0} = \Pi_{m=p}^{k} \Lambda^p_{[m]}, \tag{17}$$

where $\Lambda^p_{[m]} \triangleq p(\boldsymbol{Z}_{[m]}=\boldsymbol{z}_{[m]}|\mathrm{H}_0, \boldsymbol{Z}_{[p:m-1]}=\boldsymbol{z}_{[p:m-1]})$ comes from a MH-PDAF matched to $\mathscr{M}_1$ initialized at step $p-1=k-N$ with the moments and weights describing the posterior pdf of $\boldsymbol{X}_{[p-1]}$. As for $\mathrm{H}_1(\theta_1)$, it writes as, with $V=2T$,

$$\Lambda_{\mathrm{H}_1}(\theta_1) = \begin{cases} V^{-N} & \text{if } \theta_1 = p, \\ \Pi_{m=p}^{\theta_1-1} \Lambda^p_{[m]} \times V^{-(k-\theta_1+1)} & \text{otherwise.} \end{cases} \tag{18}$$

If the decision outcome is $\mathrm{H}_0$, then the $N$ sets of moments and weights computed at times $p, \ldots, k$ from the MH-PDAF

**Algorithm 3** The GLRTBF1.

$\Big[\{\hat{x}_{[m|m]}, P_{[m|m]}\}_{m=p,\dots,k}, \{\{W^j_{[m]}\}_{j=1,\dots,J}\}_{m=p,\dots,k}, \{\{\hat{x}^j_{[m|m]}, P^j_{[m|m]}\}_{j=1,\dots,J}\}_{m=p,\dots,k}\Big]$
$= \mathsf{GLRTBF1}\big(\{\boldsymbol{z}_{[m]}\}_{m=p,\dots,k}, \{W^j_{[p-1]}\}_{j=1,\dots,J}, \{\hat{\bar{x}}^j_{[p-1|p-1]}, \bar{P}^j_{[p-1|p-1]}\}_{j=1,\dots,J}, \{\boldsymbol{u}_{[m]}\}_{m=p-1,\dots,k-1}\big)$

**FOR** $m = p,\dots,k$ **DO**
  Using a MH-PDAF matched to $\mathscr{M}_1$, Compute the moments and weights of $\boldsymbol{X}_{[m]}$ assuming $\mathrm{H}_0$ is true, and get the likelihood of $\mathrm{H}_0$ w.r.t. $\boldsymbol{z}_{[m]}$, $\Lambda^p_{[m]} \triangleq p(\boldsymbol{Z}_{[m]} = \boldsymbol{z}_{[m]} | \mathrm{H}_0, \boldsymbol{Z}_{[p:m-1]} = \boldsymbol{z}_{[p:m-1]})$:
  $\Big[\hat{\bar{x}}_{[m|m]}, \bar{P}_{[m|m]}, \{\bar{W}^j_{[m]}\}_{j=1,\dots,J}, \{\hat{\bar{x}}^j_{[m|m]}, \bar{P}^j_{[k|k]}\}_{j=1,\dots,J}, \Lambda^p_{[m]}\Big] = \mathsf{MH\text{-}PDAF}\big(\boldsymbol{z}_{[m]}, \{\bar{W}^j_{[m-1]}\}_{j=1,\dots,J}, \{\hat{\bar{x}}^j_{[m-1|m-1]}, \bar{P}^j_{[m-1|m-1]}\}_{j=1,\dots,J}, \boldsymbol{u}_{[m-1]}\big)$
**END FOR**
Compute the likelihood of $\mathrm{H}_0$ w.r.t. $\boldsymbol{z}_{[p:k]}$: $\Lambda_{\mathrm{H}_0} = \Pi^k_{m=p} \Lambda^p_{[m]}$
**FOR** $\theta_1 = p,\dots,k$ **DO**
  Compute the likelihood of $\mathrm{H}_1(\theta_1)$ w.r.t. $\boldsymbol{z}_{[p:k]}$: $\Lambda_{\mathrm{H}_1}(\theta_1) = \begin{cases} V^{-N} & \text{if } \theta_1 = p, \\ \Pi^{\theta_1-1}_{m=p} \Lambda^p_{[m]} \times V^{-(k-\theta_1+1)} & \text{otherwise.} \end{cases}$
**END FOR**
Proceed to the GLRT: $\frac{\Lambda_{\mathrm{H}_1}(\hat{\theta}_1)}{\Lambda_{\mathrm{H}_0}} \underset{\mathrm{H}_0}{\overset{\mathrm{H}_1}{\gtrless}} \eta$, with $\hat{\theta}_1 \triangleq \mathrm{argmax}_{\theta_1 \in \mathscr{W}} \big[\Lambda_{\mathrm{H}_1}(\theta_1)\big]$.
**IF** $\mathrm{H}_0$ **detected THEN**
  keep the $N$ sets of moments and weights given by the MH-PDAF as the global output:
  $\big\{\{\hat{x}_{[m|m]}, P_{[m|m]}\}, \{W^j_{[m]}\}_{j=1,\dots,J}, \{\hat{x}^j_{[m|m]}, P^j_{[m|m]}\}_{j=1,\dots,J}\big\}_{m=p,\dots,k} = \big\{\{\hat{\bar{x}}_{[m|m]}, \bar{P}_{[m|m]}\}, \{\bar{W}^j_{[m]}\}_{j=1,\dots,J}, \{\hat{\bar{x}}^j_{[m|m]}, \bar{P}^j_{[m|m]}\}_{j=1,\dots,J}\big\}_{m=p,\dots,k}$.
**END IF**
**IF** $\mathrm{H}_1$ **detected THEN**
  keep the $\hat{\theta}_1 - 1$ first sets of moments and weights given by the MH-PDAF:
  $\big\{\{\hat{x}_{[m|m]}, P_{[m|m]}\}, \{W^j_{[m]}\}_{j=1,\dots,J}, \{\hat{x}^j_{[m|m]}, P^j_{[m|m]}\}_{j=1,\dots,J}\big\}_{m=p-1,\dots,\hat{\theta}_1-1} = \big\{\{\hat{\bar{x}}_{[m|m]}, \bar{P}_{[m|m]}\}, \{\bar{W}^j_{[m]}\}_{j=1,\dots,J}, \{\hat{\bar{x}}^j_{[m|m]}, \bar{P}^j_{[m|m]}\}_{j=1,\dots,J}\big\}_{m=p-1,\dots,\hat{\theta}_1-1}$.
  **FOR** $m = \hat{\theta}_1,\dots,k$ **DO**
    Using a MH-Pred matched to $\mathscr{M}_2$, recompute the moments and weights of $\boldsymbol{X}_{[m]}$ :
    $\Big[\hat{x}_{[m|m]}, P_{[m|m]}, \{W^j_{[m]}\}_{j=1,\dots,J}, \{\hat{x}^j_{[m|m]}, P^j_{[m|m]}\}_{j=1,\dots,J}\Big] = \mathsf{MH\text{-}Pred}\big(\{W^j_{[m-1]}\}_{j=1,\dots,J}, \{\hat{x}^j_{[m-1|m-1]}, P^j_{[m-1|m-1]}\}_{j=1,\dots,J}, \boldsymbol{u}_{[m-1]}\big)$.
  **END FOR**
**END IF**

matched to $\mathscr{M}_1$ are kept unchanged. Otherwise, if $\mathrm{H}_1(\hat{\theta}_1)$ is detected, only the first $\hat{\theta}_1 - p$ sets, at times $p,\dots,\hat{\theta}_1 - 1$, are kept unmodified. A MH-Pred matched to $\mathscr{M}_2$ is then initialized at $\hat{\theta}_1 - 1$ with the corresponding posterior moments and weights, so as to recompute their values at the subsequent $k - \hat{\theta}_1 + 1$ instants, from time $\hat{\theta}_1$ to $k$. The GLRTBF1 strategy to detect a switch $\mathscr{M}_1 \to \mathscr{M}_2$ is summarized in Algorithm 3.

*2) Detection of a switch $\mathscr{M}_2 \to \mathscr{M}_1$:* Assume that the system obeys to $\mathscr{M}_2$ before time $p \triangleq k - N + 1$. The hypotheses to be tested are

$$\begin{aligned} &\mathrm{H}_1(\theta_1): \text{ a switch } \mathscr{M}_2 \to \mathscr{M}_1 \text{ occurred at } \theta_1 \in \mathscr{W}, \\ &\mathrm{H}_0: \text{the system has remained in } \mathscr{M}_2. \end{aligned}$$

The likelihoods $\Lambda_{\mathrm{H}_0} \triangleq p(\boldsymbol{Z}_{[p:k]} = \boldsymbol{z}_{[p:k]} | H_0)$ and $\Lambda_{\mathrm{H}_1}(\theta_1) \triangleq p(\boldsymbol{Z}_{[p:k]} = \boldsymbol{z}_{[p:k]} | H_1(\theta_1))$ of $\mathrm{H}_0$ and $\mathrm{H}_1(\theta_1)$ write as

$$\Lambda_{\mathrm{H}_0} = V^{-N}, \qquad \Lambda_{\mathrm{H}_1}(\theta_1) = V^{-(\theta_1 - p)} \times \Pi^k_{m=\theta_1} \Lambda^{\theta_1}_{[m]}, \quad (19)$$

where $\Lambda^{\theta_1}_{[m]}$ comes from a MH-PDAF matched to $\mathscr{M}_1$ initialized at time $\theta_1 - 1$ with the posterior moments and weights obtained by a MH-Pred, itself initialized at $p-1$ with the (given) moments and weights describing the posterior pdf of $\boldsymbol{X}_{[p-1]}$. As the computation of $\Lambda_{\mathrm{H}_1}(\theta_1)$ for every candidate $\theta_1$ requires $N$ distinct MH-PDAFs, the complexity of the detection of a switch $\mathscr{M}_2 \to \mathscr{M}_1$, denoted GLRTBF2, becomes higher than that of GLRTBF1.

*3) Important issues:* To make the GLRT reliable, a large data window length $N$ is recommended. However, a high $N$ augments the delay in detection and compensation, increases the numerical complexity of GLRTBF2, and may disrespect the sporadic assumption on model switches. So, a tradeoff is necessary. Secondly, for large dynamic noises, the state posterior pdf quickly spreads out when the source stops uttering, which may lead to the aforementioned overconfident conclusions. Solutions may consist in approximating the state posterior pdf by a mixture of sharper pdfs and/or to reinitialize the filter after a long source pause.

## V. Experimentation

### A. Experimental setup

To assess the approach with real binaural signals, experiments were conducted in an acoustically prepared room, equipped with 3D pyramidal pattern studio foams on the roof and the walls. Two identical omnidirectional microphones, spaced by 17cm, were mounted on the top of a tripod, itself placed on a mobile wheeled cart. The two microphones outputs were synchronously acquired at $f_s = 44.1$kHz. The sensor was moved manually while the source, a loudspeaker placed at the same height, was emitting intermittent speech signal. The ground-truth source and sensor positions and velocities were determined at 200Hz by an infrared-based motion capture system with less than 1mm error.

### B. Experimental results

*1) Intermittent static source:* Figure 3 depicts the localization results brought back in the world frame at the four times $\{1, 37, 115, 275\}$. The localization runs at 20Hz, but this rate can be easily modified. At initial time, the filter handles $J = 24$ Gaussian priors (corresponding to the modes $\{F_j\}_{j=1,\dots,J}$) defined so that the union of their 99 % probability ellipsoids covers a 4m radius circular region around the sensor. At time 37, part of the remaining modes of the MH-PDAF spread along the source-sensor direction, while others extend

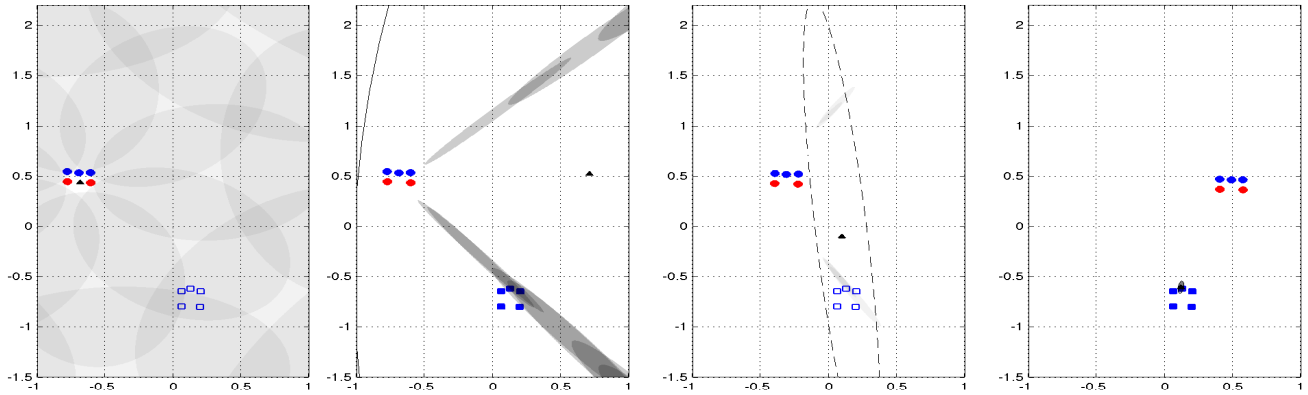
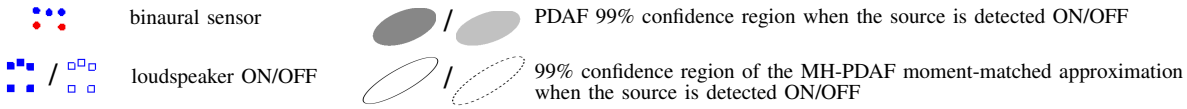


Fig. 3: Estimation results in the world frame at times $\{1, 37, 115, 275\}$ (from left to right) for a static intermittent source




along the symmetric direction w.r.t. the $(R_l R_r)$ axis. In other words, the uncertainty on the distance to the source is high, and the front-back ambiguity remains. This comes from the use of ITD cue in the filter: ITD is indeed known to carry few information about range, and cannot disambiguate sounds coming from rear and front. At iteration 115, the loudspeaker is mute, and the filter has detected the transition $\mathscr{M}_1 \rightarrow \mathscr{M}_2$ So, the state pdf is propagated along time through prediction only, *i.e.* the measurements are not incorporated into the filter. The integration of the dynamic noise characteristics along time results in a seamless growth of the 99 % probability ellipsoids. At iteration 275, the loudspeaker utters again, and the transition $\mathscr{M}_2 \rightarrow \mathscr{M}_1$ has been detected. Thanks to the information brought by motion, front and back have been disambiguated and the range uncertainty has been lowered, which leads to a state pdf very sharp around the true source location. The companion video (see also `homepages.laas.fr/danes/IROS2012`) shows the extracted ITD and the detected source activity along time.

*2) Dynamic source:* Recall that if the velocity vector $\boldsymbol{U_2}$ is unknown, then (1) has to be complemented to account for its prior dynamics (Section II). In this experiment, $\boldsymbol{U_2}$ is assumed to follow a random walk. As shown on the companion video, the posterior confidence region is wider than in the above static case, especially in the emitter-to-receiver direction. This is so because additional uncertainty is put on the source motion. More generally, the filter usefulness depends on whether the source dynamic model is permissive or not (*i.e.* if much information is available about the source motion or not).

## VI. Conclusion

A stochastic estimation strategy has been proposed to the localization of a single mobile source. By fusing the sensor motion with its perception, both the source range and azimuth can be estimated, and front-back positions can be disambiguated. The strategy copes with false measurements, and handles source intermittency. Experimental results show the effectiveness of the approach. Ongoing extensions aim to deal with multiple sources, scattering effects induced by a head between the microphones, and room reverberations.

## References


[1] D. Rosenthal and H. Okuno, *Computational Auditory Scene Analysis*. Lawrence Erlbaum Associates, 1998.

[2] K. Nakadai, H. Okuno, and H. Kitano, "Epipolar geometry based sound localization and extraction for humanoid audition," in *IEEE Int. Conf. on Intell. Robots and Systems (IROS'2011)*, 2011.

[3] J. 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, 2007.

[4] A. Laflaquiere, S. Argentieri, B. Gas, and E. Castillo-Castaneda, "Space dimension perception from the multimodal sensorimotor flow of a naive robotic agent," in *IEEE/RSJ Int. Conf. on Intell. Robots and Systems (IROS'10)*, Taipei, Taiwan, 2010.

[5] K. Nakadai, T. Lourens, H. Okuno, and H. Kitano, "Active audition for humanoid," in *Nat. Conf. on Artificial Intelligence*, 2000.

[6] I. Markovic and I. Petrovic, "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.

[7] Y.-C. Lu and M. Cooke, "Motion strategies for binaural localisation of speech sources in azimuth and distance by artificial listeners," *Speech Communication*, 2010.

[8] A. Portello, P. Danès, and S. Argentieri, "Acoustic models and Kalman filtering strategies for active binaural sound localization," in *IEEE/RSJ Int. Conf. on Intell. Robots and Systems (IROS'2011)*, 2011.

[9] A. Portello, P. Danès, S. Argentieri, and M. Kumon, "Active binaural localization of a moving speech source," 2012, available on request.

[10] H. Viste and G. Evangelista, "Binaural source localization," in *Int. Conf. on Digital Audio Effects (DAFx'04)*, 2004.

[11] C. Knapp and G. Carter, "The generalized correlation method for estimation of time delay," *IEEE Trans. on Acoustics, Speech, and Signal Processing*, 1976.

[12] E. Weinstein and A. Weiss, "Fundamental limitations in passive time delay estimation - Part II: Wideband systems," *IEEE Trans. on Acoustics, Speech and Signal Processing*, 1984.

[13] A. Weiss and E. Weinstein, "Fundamental limitations in passive time delay estimation - Part I: Narrowband systems," *IEEE Trans. on Acoustics, Speech and Signal Processing*, 1983.

[14] S. Julier and J. Uhlmann, "Unscented filtering and nonlinear estimation," *Proc. of IEEE*, 2004.

[15] R. Van der Merwe and E. Wan, "The square-root unscented kalman filter for state and parameter estimation," in *IEEE Int. Conf. on Acoustics, Speech and Signal Processing (ICASSP'01)*, 2001.

[16] Y. Bar-Shalom and X. Li, *Multitarget-Multisensor Tracking: Principles and Techniques*. YBS Publishing, 1995.

[17] A. Willsky, "Detection of abrupt changes in dynamic systems," in *Detection of Abrupt Changes in Signals and Dynamical Systems*, ser. LNCIS. Springer-Verlag, 1986, no. 77, pp. 27–49.