Multi-Step-Ahead Information-Based Feedback Control for Active Binaural Localization

Gabriel Bustamante, Patrick Danès

Multi-Step-Ahead Information-Based Feedback Control for Active Binaural Localization

Gabriel Bustamante[1] and Patrick Danès[1]

***Abstract*— Binaural sound localization is known to be improved by incorporating the movement of the sensor. "Active" schemes based on this paradigm can overcome conventional limitations such as front-back ambiguity and source range recovery. Starting from a Gaussian prior on the relative position of a source, this paper determines the motion of a binaural sensor which leads to the most effective path for localization. To this aim, a reward function is defined as the conditional expectation, over the yet unknown N next observations, of the entropy of the N-step-ahead posterior pdf of the relative source position. The optimal motion of the binaural sensor is obtained from a constrained optimization problem involving the automatic differentiation of the reward function. The method is validated in simulation, and is being implemented on a real-life robotic test bed.**

I. INTRODUCTION

In robot audition, binaural localization of sound sources has been improved by the advent of "active" schemes, combining binaural perception and motor commands. With such techniques, the front-back ambiguity can be eliminated and the source range can be recovered [1]. Then, the exploration problem naturally arises, *i.e.*, how to drive the binaural head so as to improve the efficiency of the localization process.

This is related to information-theoretic control problems which consist in determining system inputs in order to maximize an information criterion one or several steps ahead. In robotics, Simultaneous Localization and Mapping (SLAM) techniques have been extended to cope with this objective [2]. For instance, control inputs to a mobile robot can be determined with a single-step-ahead method so as to improve the knowledge of the environment, this information being expressed in terms of Shannon entropy [3]. As a sequence of one-step-ahead optimal controls does not necessary lead to the best information [4] at the end of a time horizon, multi-step methods have been investigated. The control of a robot-mounted camera to optimize depth estimation has been addressed [5], by maximizing an information gain several steps ahead. Therein, future unknown observations are assumed to match their predicted mean, as in [6]. Using the same assumption, a multi-step ahead entropic criterion has been proposed [7] to select the optimal zoom of a motor-controlled camera in an object tracking task.

In robot audition, an audio-based motion planning strategy has been proposed to improve speech recognition with one microphone [8]. In [9], an agent selects pre-defined actions through (time-consuming) Monte Carlo Exploration in order to approach a specific goal while reducing the entropy of the one-step-ahead belief on a source position. An approximate but tractable multi-step-ahead approach has been set up to improve audio source localization by a robot equipped with a microphone array [10]. The belief of the source position is represented on a discrete grid, then optimal robot commands over a fixed horizon are computed to minimize the expected entropy of the grid. In [11], an optimal long-term robot motion planning algorithm is proposed for active source localization by performing (time-consuming) Monte Carlo tree search.

This paper develops a multi-step-ahead control strategy for sound source localization. First, the audio model is presented, consisting in a rigid-body dynamics along with an explicit measurement equation depicting audio cues for exploration. Then the definition and computation of the exploration reward function is addressed. The gradient of this function is computed by means of automatic differentiation and used in a constrained optimization problem leading to the solution. Finally, the method is evaluated through simulations on realistic data. Some aspects related to its ongoing experimental assessment are discussed.

II. PROBLEM STATEMENT

A. Audio-motor model

Consider a mobile robot equipped with a binaural sensor, *i.e.*, two microphones R_1 and R_2 laid on an anthropomorphic head. The position of the binaural sensor at discrete time k is denoted by the frame $\mathcal{F}_k = (O_k, \overrightarrow{x_{Rk}}, \overrightarrow{y_{Rk}}, \overrightarrow{z_{Rk}})$ with O_k the center of $\|\overrightarrow{R_1R_2}\|$. The vector $\overrightarrow{y_{Rk}} = \frac{\overrightarrow{R_2R_1}}{\|\overrightarrow{R_2R_1}\|}$ supports the interaural axis while $\overrightarrow{z_{Rk}}$ points frontwards. An omnidirectional sound source E emits continously. It is supposed to lie on the plane defined by $(O_k, \overrightarrow{y_{Rk}}, \overrightarrow{z_{Rk}})$, and its relative coordinates in $\mathcal{F}_k$ are denoted by $x_k = (0, e_{yk}, e_{zk})^T$. When applying the control input vector to the head, the head-to-source position varies according to the stochastic discrete-time state equation

$$x_{k+1} = f(x_k, u_k) + w_k,\ \ w_k \sim \mathcal{N}(0, Q_k), \tag{1}$$

with w_k a Gaussian zero-mean white dynamic noise of covariance matrix Q_k.

The short-term analysis of the binaural stream leads to the extraction of binaural cues such as interaural differences or azimuth likelihood [12]. At time k, a posterior probability density function (pdf – or "belief") $p(x_k|y_{1:k})$ of the head-to-source position is assumed to be available, *e.g.*, on

This work was partially supported by EU FET grant TWO!EARS, ICT-618075, www.twoears.eu

[1] LAAS-CNRS, Université de Toulouse, CNRS, INPT, UPS, Toulouse, France. Tel: +33 (0)5 61 33 78 25, Fax: +33 (0)5 61 33 64 55, E-mail: patrick.danes@laas.fr

the basis of cues $y_{1:k} = y_1, \ldots, y_k$ and motor commands $u_{0:k-1} = u_0, \ldots, u_{k-1}$. For instance, the Gaussian mixture square-root unscented Kalman filter (GMsrUKF) proposed in [13] can be used. This solution incorporates the likelihoods $p(y_k|\theta_k)$ of the source azimuths θ_k along time, so as to express $p(x_k|y_{1:k})$ as a gaussian mixture.

B. N-step ahead exploration statement

The aim of the exploration strategy is to compute from the belief at time k, the N next control inputs $\bar{u}_N = u_{k:k+N-1}$ of the binaural head which lead, on average, to the "best" belief at the end of the N-step horizon, *i.e.*, at time $k+N$. This belief depends on the yet unknown observation variables $y_{k+1:k+N}$. To guide the exploration, alternative future observation variables $z_{k+1:k+N}$ (which can differ from $y_{k+1:k+N}$) are defined. In the proposed approach, these must be related to the hidden state vector x_k and azimuth $\theta_k = -\mathrm{atan2}(e_y, e_z)$ by a scalar explicit observation model

$$z_k = h(x_k) + v_k = \bar{h}(\theta_k) + v_k,\ \ v_k \sim \mathcal{N}(0, R_k), \qquad (2)$$

with v_k the scalar Gaussian white zero-mean measurement noise of variance R_k. Importantly, R_k must not depend on the hidden value of x_k. For instance, when the microphones are laid on a spherical head and when the source is farfield, $\bar{h}(\theta_k)$ can express the Woodworth-Schlosberg approximation of the interaural time difference (ITD) [14].

When the motion of the sensor is to be controlled in real time, a predictive control strategy can be implemented, which consists in applying the first element $u_k^\star$ of the N-step-ahead optimal solution $\bar{u}_N^\star$. Then, the new measurement y_{k+1} becomes available. Its assimilation leads to the new belief $p(x_{k+1}|y_{1:k+1})$ at time $k+1$, and the overall process can be repeated. The question thus arises of the influence of the selection of N on the obtained behavior, *e.g.*: to which extent are the benefits of a N-step-ahead strategy (computed from the average behavior of the system w.r.t. $z_{k+1} : z_{k+N}$) better than those of a single-step ahead optimum control?

C. Simplifying assumptions

To develop the exploration strategy, some simplifying assumptions are made. First, at time k, a Gaussian belief with mean $\hat{x}_{k|k}$ and covariance $P_{k|k}$, denoted by $p(x_k|z_{1:k}) = \mathcal{N}(x_k; \hat{x}_{k|k}, P_{k|k})$ is defined to initialize the exploration strategy. It can typically be deduced from the gaussian mixture $p(x_k|y_{1:k})$ by keeping its hypothesis of maximum weight, by computing its moment-matched approximation, etc. Second, by denoting $R(\phi)$ the rotation matrix of angle ϕ, the function f in (1) is defined as

$$\begin{aligned} f : \mathbb{R}^3 &\to \mathbb{R} \\ x_k &\mapsto x_{k+1} = R^T(\phi_k)x_k - R^T(\phi_k)T_k, \end{aligned} \qquad (3)$$

i.e., $\mathcal{F}_{k+1}$ is the image of $\mathcal{F}_k$ by the rigid transform of 2D translation vector $T_k \triangleq \begin{pmatrix} \overrightarrow{O_kO_{k+1}}.\overrightarrow{y_{R_k}} \\ \overrightarrow{O_kO_{k+1}}.\overrightarrow{z_{R_k}} \end{pmatrix}$ and rotation angle $\phi_k \triangleq (\widehat{\overrightarrow{z_{R_k}}, \overrightarrow{z_{R_{k+1}}}})_{\text{around}\overrightarrow{x_{R_k}}}$, see Figure 1. Therefore, the control vector at time k is defined as $u_k = (T_k, \phi_k)^T$. The dynamic noise covariance Q_k is supposed small enough

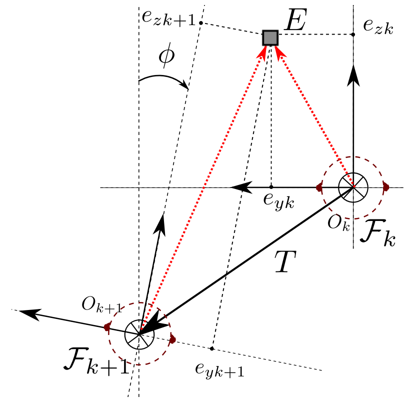


Fig. 1: Rigid motion of the binaural sensor

so that at first approximation, the loss of information during the movement can be neglected. Indeed, if the binaural sensor undergoes a noiseless rigid motion, then $p(x_k|z_{1:k})$ and $p(x_{k+1}|z_{1:k})$ describe the same reality but expressed from the viewpoints of frames $\mathcal{F}_k$ and $\mathcal{F}_{k+1}$, respectively.

III. FEEDBACK CONTROL FOR BINAURAL LOCALIZATION

A. Multi-step ahead information-theoretic reward function

The differential entropy of the stochastic vector x_k is a measure of uncertainty of x_k. If x_k is distributed according to the conditional pdf $p(x_k|z_{1:k})$, then its entropy is defined as $H(x_k|z_{1:k}) = -\int p(x_k|z_{1:k}) \log p(x_k|z_{1:k}) dx$. Likewise, the mutual information of two random vectors x_k, z_k jointly distributed according to $p(x_k, z_k|z_{1:k-1})$ is equal to $I(x_k, z_k|z_{1:k-1}) = I_k$ with $I_k = \int p(x_k, z_k|z_{1:k-1}) \log \frac{p(x_k, z_k|z_{1:k-1})}{p(z_k|z_{1:k-1})p(x_k|z_{1:k-1})} dx_k dz_k$. $I(x_k, z_k|z_{1:k-1})$ embodies the amount of information that x_k contains about z_k [15]. The entropy of a Gaussian random vector is an increasing function of the log determinant of its covariance matrix, *e.g.*, if $p(x_k|z_{1:k}) = \mathcal{N}(x_k; \hat{x}_{k|k}, P_{k|k})$ then

$$H(x_k|z_{1:k}) = \frac{1}{2}\log((2\pi e)^2 \det(P_{k|k})). \qquad (4)$$

From the belief $p(x_k|z_{1:k})$ on the source position at time k, the objective is to compute the sequence of control inputs $\bar{u}_N$ that minimizes a reward function J_N defined from the entropy $H(x_{k+N}|z_{1:k+N})$. Since this entropy depends on the yet unknown N future observations $z_{k+1:k+N}$, J_N is set to the conditional expectation

$$J_N = \mathbb{E}_{z_{k+1:k+N}|z_{1:k}}\left[H(x_{k+N}|z_{1:k+N})\right]. \qquad (5)$$

From the assumptions defined in (II-C) and the Bayes rule, the one-step-ahead reward function J_1 can be set up as

$$\begin{aligned} J_1 &= K_1' - H(z_{k+1}|z_{1:k}) \\ &= K_1' - F_1(\bar{u}_1), \end{aligned} \qquad (6)$$

with $\bar{u}_1 = u_k$ the next control input vector, K_1' a scalar constant independent of $\bar{u}_1$, and $H(z_{k+1}|z_{1:k}) = F_1(\bar{u}_1)$ the entropy associated to the next predicted measurement pdf

$p(z_{k+1}|z_{1:k})$, which depends on $\bar{u}_1$. For $N \geq 2$, the reward function $J_N(\bar{u}_N)$ comes as

$$\begin{aligned} J_N &= K'_N - H(z_{k+1}|z_{1:k}) \\ &\quad - \sum_{i=2}^{N} \mathbb{E}_{z_{k+1:k+i-1}|z_{1:k}} \left[H(z_{k+i}|z_{1:k+i-1}) \right] \\ &= K'_N - F_1(\bar{u}_1) \\ &\quad - \sum_{i=2}^{N} \mathbb{E}_{z_{k+1:k+i-1}|z_{1:k}} \left[F_i(\bar{u}_i, z_{k+1:k+i-1}) \right], \end{aligned} \tag{7}$$

where K'_N is a scalar constant independent of $\bar{u}_N$, $H(z_{k+i}|z_{1:k+i-1})$ depends on the sequence $\bar{u}_i = u_{k:k+i-1}$ of control inputs up to time $k+i-1$ and on the future observations $z_{k+1:k+i-1}$, and is thus termed $F_i(\bar{u}_i, z_{k+1:k+i-1})$. The proofs of equations (6) and (7) are given in appendix.

Since the function F_i cannot be expressed in closed-form, a difficulty arises in the computation of the expectation integral in (7). For $i \geq 2$, the conditional pdf $p(z_{k+1}, \ldots, z_{k+i-1}|z_{1:k})$ of the random vector $z_{k+1:k+i-1}$ (of size $i-1$) is approximated by the Gaussian pdf $\mathcal{N}(z_{k+1:k+i-1}; \hat{z}_{k+1:k+i-1|k}, C)$ centered on $\hat{z}_{k+1:k+i-1|k}$ with covariance C, by means of the unscented transform. On this basis, $\mathbb{E}_{z_{k+1:k+i-1}|z_{1:k}} \left[F_i(\bar{u}_i, z_{k+1:k+i-1}) \right]$ is approximated from the evaluation of F_i at the $2(i-1)+1$ sigma-points $\{Z_j\}$, by a linear combination involving the weights $\{W_j\}$ of the unscented transform, namely,

$$\int F_i(\bar{u}_i, z_{k+1:k+i-1}) p(z_{k+1}, \ldots, z_{k+i-1}|z_{1:k}) dz_{k+1:k+i-1} \approx \sum_{j=1}^{2(i-1)+1} W_j F_i(\bar{u}_i, Z_j), \tag{8}$$

in the vein of [16]. The sigma-points $\{Z_j\}$ are deterministically drawn from the Gaussian approximation $\mathcal{N}(z_{k+1:k+i-1}; \hat{z}_{k+1:k+i-1|k}, C)$ of $p(z_{k+1}, \ldots, z_{k+i-1}|z_{1:k})$, what involves the Cholesky decomposition $C = LL^T$. Finally, the reward function J_N can be rewritten as

$$J_N = K'_N - F_1(\bar{u}_1) - \sum_{i=2}^{N} \sum_{j=1}^{2(i-1)+1} W_j F_i(\bar{u}_i, Z_j). \tag{9}$$

B. Gradient of the reward function

Computing the gradient of the reward function J_N w.r.t. the vector made up with the control input sequence $\bar{u}_N$ is crucial to study the variations of the information with the movement and to set up an optimization problem. Denote the gradient operator as $\nabla_{\bar{u}_N} = (\nabla^T_{u_k}, \nabla^T_{u_{k+1}}, \ldots, \nabla^T_{u_{k+N-1}})^T$. In the expression

$$\nabla_{\bar{u}_N} J_N = -\nabla_{\bar{u}_N} F_1(\bar{u}_1) - \sum_{i=2}^{N} \sum_{j=1}^{2(i-1)+1} W_j \nabla_{\bar{u}_N} F_i(\bar{u}_i, Z_j), \tag{10}$$

$F_i(\bar{u}_i, Z_j)$ does not depend on control inputs applied after time $k+i-1$. So, $\nabla_{u_{k+i}} F_i = \ldots = \nabla_{u_{k+N-1}} F_i = 0$, and $\nabla_{\bar{u}_N} F_i$ solely depends on $\{\nabla_{\bar{u}_l} F_l\}_{l \leq i}$. As F_i has no closed-form equation, its gradient cannot be evaluated straightforwardly. To avoid finite differences methods which arise the difficulty of finding the balance between numerical precision and truncation errors, a forward accumulation automatic differentiation algorithm has been implemented [17]. The program that computes F_i for specific control inputs values $\bar{u}_i = \bar{u}_{i0}$ is complemented by automatic differentiation, so as to compute the gradient $\nabla_{\bar{u}_i} F_i(\bar{u}_{i0}, z_{k+1:k+i-1})$ w.r.t. $\bar{u}_i$ at these values. The algorithm relies on dual numbers algebra, which extends the set of real numbers by adding a nilpotent element ε such that $\varepsilon^2 = 0$. So, any dual number z_d writes as $z_d = z + \varepsilon \dot{z}$, with $(z, \dot{z})$ a pair of (real) value and (real) derivative (*e.g.*, the value and derivative of a given function at a given point). For instance, suppose that $a_d = a + \varepsilon \dot{a}$, $b_d = b + \varepsilon \dot{b}$. Then the dual number $z_d = b_d \cos(a_d)$ writes as $z + \varepsilon \dot{z}$ with

$$z = b \cos(a), \tag{11}$$

$$\dot{z} = \dot{b} \cos(a) - \dot{a} b \sin(a), \tag{12}$$

which respectively correspond to the value and derivative of $g \cos(f)$ at $f = a, \dot{f} = \dot{a}, g = b, \dot{g} = \dot{b}$.

C. Constrained optimization problem

From the belief $p(x_k|z_{1:k})$ of the source position at time k, the reward function J_N (9) has been set up. While minimizing J_N is crucial, feasibility of the control input sequence (*i.e.*, limitations due to the motion capacities) need to be taken into account. The sets $\mathcal{T} = \{(T_y, T_z) \in \mathbb{R}^2 \mid {T_y}^2 + {T_z}^2 \leq r^2_{max}\}$ and $\mathcal{R} = \{\phi \in \mathbb{R} \mid |\phi| \leq \phi_{max}\}$ define the admissible translation and rotation, with r_{max} (resp. ϕ_{max}) the maximum distance reachable by the robot (resp. the maximum possible rotation of the head) between two consecutive time steps. The constrained optimization problem $\mathcal{P}_N$ follows:

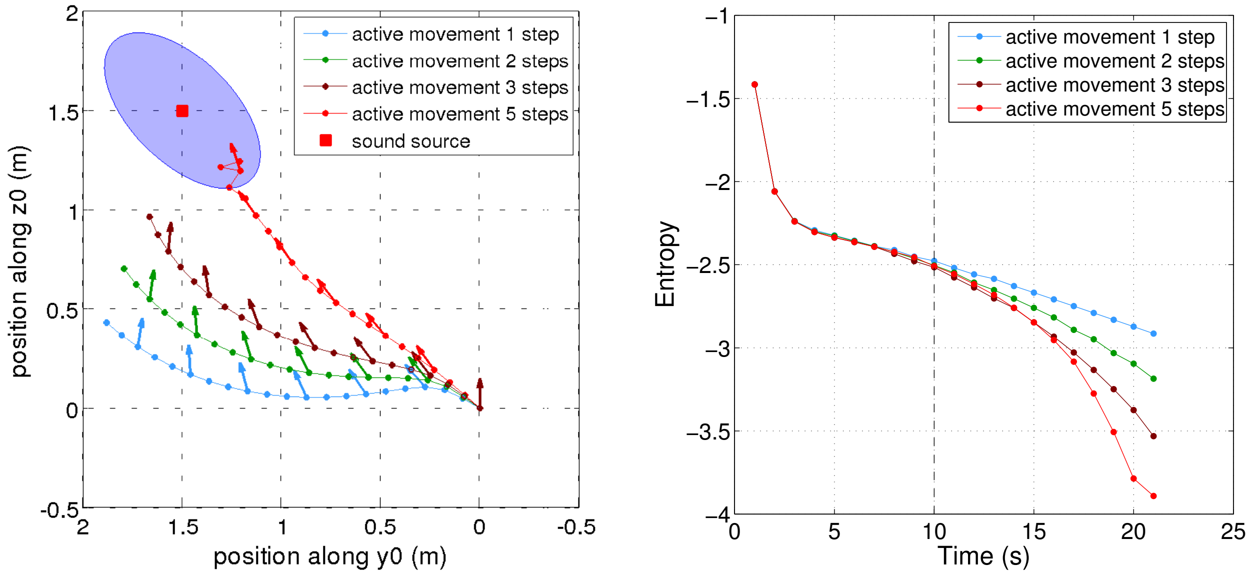
$$(\mathcal{P}_N) \begin{cases} \bar{u}^\star_N &= \arg\min_{\bar{u}_N \in (\mathcal{T}\times\mathcal{R})^N} J_N(\bar{u}_N) \\ &= \arg\max_{\bar{u}_N \in (\mathcal{T}\times\mathcal{R})^N} F_1(\bar{u}_1) + \sum_{i=2}^{N} \sum_{j=1}^{2(i-1)+1} W_j F_i(\bar{u}_i, Z_j) \end{cases} \tag{13}$$

Using the gradient (10), pointing to the direction of steepest ascent of J_N, the problem $\mathcal{P}_N$ is numerically solved by means of a projected gradient algorithm. The values of each element u_{k+i} (for $i \in \{0, \ldots, N-1\}$) of the general control input vector $\bar{u}_N$, are iteratively updated through the conventional gradient ascent method, then projected onto the closed convex set $(\mathcal{T} \times \mathcal{R})$ by the operator

$$\pi_{\mathcal{T}\times\mathcal{R}}(u) \triangleq \arg\min_x \left\{ \|u - x\|_2, x \in (\mathcal{T} \times \mathcal{R}) \right\}. \tag{14}$$

IV. Simulations with realistic data

The whole binaural localization framework including, the short-term detection of azimuth and the multi-Gaussian filtering strategy has been implemented on a simulated KEMAR binaural head-and-torso-simulator (HATS) from G.R.A.S.®(`kemar.us`). The binaural head is supposed to



(a) (b)

Fig. 2: Simulated sound source localization for different exploration horizons. The control inputs are computed from the minimization of the reward function J_1,J_2,J_3 or J_5 at each sampling time. (a): Source position, localization of the sound source with the 99%-probability confidence ellispoids of the initial belief and head trajectories in the world frame (*i.e.*, in the initial frame $\mathcal{F}_0$) . (b): Decrease of the entropy of the filtered (posterior) state pdf for the various motion strategies. See the url at the end for further experiments.

be endowed with omnidirectional planar motion, *i.e.*, with two translational and one rotational degrees of freedom. The sound source is a non-intermittent white noise signal, filtered by a 1kHz bandwidth band-pass filter with 1kHz central frequency, as it endows the azimuth pseudo-likelihood with modes much sharper than with speech sources for instance [12]. The movements of the binaural sensor have been limited in translation and rotation by $r \leq r_{max} = 0.1$m and $|\phi| \leq \phi_{max} = 15°$.

The binaural signals perceived from the microphones have been generated online, without reverberations. While the sensor is moving, guided by the exploration strategy, those binaural signals are synthesized by using a database of Head Related Impulse Responses (HRIRs) suited to be used with KEMAR HATS and recorded in an anechoic environment. This database as well as a binaural simulator are publicly available at the URLs `www.twoears.eu` and `docs.twoears.eu/en/latest/binsim`. The sound source is initialized at the position $x_0 = (0, 1.5, 1.5)^T$ in the robot frame $\mathcal{F}_0$ at time $k = 0$. To simplify the notation in the legends of the next plots, this frame writes as $\mathcal{F}_0 = (O, \vec{x}_0, \vec{y}_0, \vec{z}_0)$.

The exploration strategy has been tested for different horizons (figure 2). In each scenario, a receding horizon exploration strategy is applied. At each time of the simulation, the sequence of control inputs is calculated from the solution of $\mathcal{P}_N$ with $N = \{1, 2, 3, 5\}$, and only its first element is applied. The entropy $H(x_k|z_{1:k})$ of the posterior belief has been evaluated for the different strategies. During the 10 first seconds, although each strategy leads to a distinct position of the binaural sensor, the four entropies are similar. The 5-step strategy drives the binaural sensor directly to the sound source, and offers the best results after $t = 10s$. This is in accordance with the intuition that driving and heading the binaural sensor towards the sound source improves the localization [4].

The entropy of the posterior pdf has also been evaluated for a large number of Monte Carlo runs for various horizon lengths $N = \{1, 2, 7\}$ (Figure 3). For each N, the computation of the first control input u_0 (applied from $k = \text{INIT}$ to $k = 1$) does not depend on the run, as it is independent of the next measurement z_1. So, given N, no spreading is observed on the entropy of the filtered state pdf at $k = 1$. Incidentally, the obtained values are close to each other for all N. At the end of an horizon defined by $N = 7$, the 7-step strategy leads to the lowest average entropy, as expected.

V. CONCLUSION

We presented a multi-step-ahead information based feedback control for binaural localization. The main contributions of this article are the theoretical reward function and the implementation of an automatic differentiation design so as to compute its gradient. The one-step-ahead exploration strategy has been implemented on a real KEMAR binaural head-and-torso endowed with omnidirectional planar motion `http://homepages.laas.fr/danes/IR2017`. The proposed multi-step-ahead strategy is being implemented too. The single-step strategy can be computed online at each time

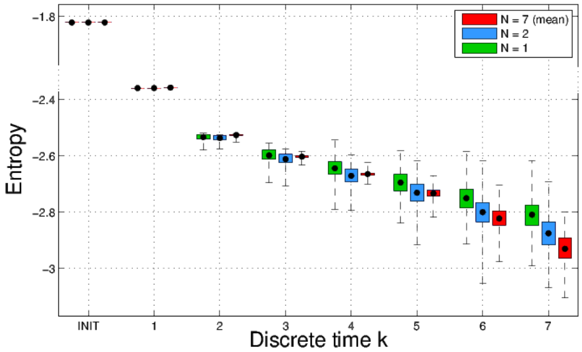


Fig. 3: Distribution of entropies of the filtered state pdfs for N-step-ahead controls, $N = \{1, 2, 7\}$. A set of 1000 experiments has been simulated for each N. In each set, the measurement z_{k+1} subsequent to the filtered state pdf at time k, has been drawn following the measurement process (2). For $N = \{1, 2\}$, the control input vector $\bar{u}_N$ is calculated at each time step and only the first element is applied. For $N = 7$, the control inputs vector is calculated once, at the initial step, and is applied throughout the horizon. The bottom and top of the boxes respectively depict the first and third quartiles, so that the boxes collect 50% of the data. The extremities of the dash lines are the extremum values. Finally, the empirical means of entropies are annotated with black dots.

of the simulation. In view of the limited computation time of the MATLAB implementation of the multi-step algorithm for small N, *i.e.*, less than 1 second for $N \leq 3$, the method is expected to work in real time. Besides, as many tasks do not only aim at localization, another prospect is to develop multiobjective feedback control policies which optimize a tradeoff between localization and a conflicting goal, such as reaching a specific position.

APPENDIX

The objective of this part is to demonstrate the results (6) and (7). First, the mutual information $I(x_{k+i}, z_{k+i}|z_{1:k+i-1})$ of the state and observation vector at time $k+i$ will be noted I_{k+i} so as to simplify the notations. Then, the following lemmata are introduced

Lemma 1: The expectation of the entropy $H(x_{k+i}|z_{1:k+i})$ of the filtered state pdf at time $k+i$ over the measurement z_{k+i} conditioned on the previous measurements can be decomposed into

$$\mathbb{E}_{z_{k+i}|z_{1:k+i-1}}\left[H(x_{k+i}|z_{1:k+i})\right] = H(x_{k+i}|z_{1:k+i-1}) - I_{k+i}. \tag{15}$$

Proof: At any time k, the entropy $H(x_{k+1}|z_{1:k+1})$ of the next filtered state pdf can easily be shown to be connected to the entropy of the next predicted state pdf and the mutual information by an update rule such as [18],[19]

$$\mathbb{E}_{z_{k+1}|z_{1:k}}\left[H(x_{k+1}|z_{1:k+1})\right] = H(x_{k+1}|z_{1:k}) - I_{k+1}. \tag{16}$$

The same relation holds between any set of consecutive time instants $k+i-1, k+i$ by direct mathematical induction, what leads to (15). ■

Lemma 2: The mutual information I_{k+i} of the state x_{k+i} and observation z_{k+i} conditioned on past measurements can be decomposed into

$$I_{k+i} = K_i + H(z_{k+i}|z_{1:k+i-1}), \tag{17}$$

where K_i is independent of the sequence of control inputs $\bar{u}_i = u_{k:k+i-1}$.

Proof: The mutual information can be decomposed into $I_{k+i} = A + B$ such that

$$A = \int \log \frac{p(x_{k+i}, z_{k+i}|z_{1:k+i-1})}{p(x_{k+i}|z_{1:k+i-1})} p(x_{k+i}, z_{k+i}|z_{1:k+i-1}) dx_{k+i} dz_{k+i}, \tag{18}$$

and

$$B = -\int \log p(z_{k+i}|z_{1:k+i-1}) p(x_{k+i}, z_{k+i}|z_{1:k+i-1}) dx_{k+i} dz_{k+i}. \tag{19}$$

The quantity A can be rewritten as

$$\begin{aligned} A &= \int \log \frac{p(x_{k+i}|z_{1:k+i-1}) p(z_{k+i}|x_{k+i})}{p(x_{k+i}|z_{1:k+i-1})} \quad (20) \\ &\qquad p(x_{k+i}|z_{1:k+i-1}) p(z_{k+i}|x_{k+i}) dx_{k+i} dz_{k+i} \\ &= \int \log p(z_{k+i}|x_{k+i}) \\ &\qquad p(x_{k+i}|z_{1:k+i-1}) p(z_{k+i}|x_{k+i}) dx_{k+i} dz_{k+i} \\ &= -\mathbb{E}_{x_{k+i}|z_{1:k+i-1}} H(z_{k+i}|x_{k+i}). \quad (21) \end{aligned}$$

Since $H(z_{k+i}|x_{k+i})$ solely depends on the covariance matrix of the measurement noise R_{k+i}, A is a constant, accordingly renamed K_i, which does not depend on $\bar{u}_i = u_{k:k+i-1}$.

$$\begin{aligned} B &= -\int \log p(z_{k+i}|z_{1:k+i-1}) \\ &\qquad p(z_{k+i}|z_{1:k+i-1}) \left(\int p(x_{k+i}|z_{1:k+i}) dx_{k+i} \right) dz_{k+i} \\ &= -\int p(z_{k+i}|z_{1:k+i-1}) \log p(z_{k+i}|z_{1:k+i-1}) dz_{k+i} \\ &= H(z_{k+i}|z_{1:k+i-1}). \quad (22) \end{aligned}$$

■

Proof: [Proof of equation (6)] Applying Lemma 1 at $i = 1$ and combining it with Lemma 2, the reward function J_1 writes as

$$J_1 = -K_1 - H(z_{k+1}|z_{1:k}) + H(x_{k+1}|z_{1:k}). \tag{23}$$

Importantly, $H(x_{k+1}|z_{1:k})$ does not depend on $\bar{u}_1 = u_k$. Indeed, it solely depends on the log-determinant of the predicted covariance matrix $P_{k+1|k} = R^T(\phi_k) P_{k|k} R(\phi_k)$, *i.e.*, in view of the properties of rotation matrices, of the log-determinant of $P_{k|k}$. Therefore, together with $-K_1$, it can be casted into a constant K_1'. Then, by denoting $H(z_{k+1}|z_{1:k})$ as the function F_1 of the control inputs $\bar{u}_1$, one gets (6). ■

Proof: [Proof of equation (7)] The demonstration is obtained by means of mathematical induction.

Basis: First, equation (7) is confirmed for $N = 2$. Considering the vector of commands $\bar{u}_2 = u_{k:k+1}$, the following holds:

$$J_2(\bar{u}_2) = \mathbb{E}_{z_{k+1:k+2}|z_{1:k}}\left[H(x_{k+2}|z_{1:k+2})\right] \\ = \mathbb{E}_{z_{k+1}|z_{1:k}}\left[\mathbb{E}_{z_{k+2}|z_{1:k+1}}\left[H(x_{k+2}|z_{1:k+2})\right]\right]. \tag{24}$$

By applying Lemma 1 at $i = 2$, J_2 becomes

$$J_2(\bar{u}_2) = \mathbb{E}_{z_{k+1}|z_{1:k}}\left[H(x_{k+2}|z_{1:k+1}) - I_{k+2}\right]. \tag{25}$$

The assumption (used in the above proof) that there is no loss of information during the head motion implies that $H(x_{k+2}|z_{1:k+1}) = H(x_{k+1}|z_{1:k+1})$. Then by (6), one gets

$$J_2(\bar{u}_2) = K_1' - F_1(\bar{u}_1) - \mathbb{E}_{z_{k+1}|z_{1:k}}\left[I_{k+2}\right] \tag{26}$$

From Lemma (2) at $i = 2$, the mutual information I_{k+2} can be replaced by $K_2 + H(z_{k+2}|z_{1:k+1})$, with K_2 constant and $H(z_{k+2}|z_{1:k+1})$ denoted by $F_2(\bar{u}_2, z_{k+1})$. Assembling the constants into K_2' lead to (7), *i.e.*,

$$J_2(\bar{u}_2) = K_2' - F_1(\bar{u}_1) - \mathbb{E}_{z_{k+1}|z_{1:k}}\left[F_2(\bar{u}_2, z_{k+1})\right]. \tag{27}$$

Inductive step: Assuming that equation (7) is true for step N, the fact that it holds for $N + 1$ is demonstrated as follows. Applying the chain rule of expectations together with Lemma 1 at $i = N + 1$ to $J_{N+1}(\bar{u}_{N+1})$ leads to

$$J_{N+1}(\bar{u}_{N+1}) = \mathbb{E}_{z_{k+1:k+N+1}|z_{1:k}}\left[H(x_{k+N+1}|z_{1:k+N+1})\right] \\ = \mathbb{E}_{z_{k+1:k+N}|z_{1:k}}\left[H(x_{k+N+1}|z_{1:k+N})\right] \\ - \mathbb{E}_{z_{k+1:k+N}|z_{1:k}}\left[I_{k+N+1}\right].$$

As $H(x_{k+N+1}|z_{1:k+N}) = H(x_{k+N}|z_{1:k+N})$, it follows that

$$J_{N+1}(\bar{u}_{N+1}) = J_N(\bar{u}_N) - \mathbb{E}_{z_{k+1}:z_{k+N}|z_{1:k+N}}\left[I_{k+N+1}\right] \\ = K_N' - F_1(\bar{u}_1) \\ - \sum_{i=2}^{N} \mathbb{E}_{z_{k+1}:z_{k+i-1}|z_{1:k}}\left[F_i(\bar{u}_i, z_{k+1:k+i-1})\right] \\ - \mathbb{E}_{z_{k+1:k+N}|z_{1:k}}\left[I_{k+N+1}\right]. \tag{28}$$

From Lemma 2 at $i = N+1$, the mutual information I_{k+N+1} can be replaced by the sum $K_{N+1} + H(z_{k+N+1}|z_{1:k+N})$, with K_{N+1} constant and $H(z_{k+N+1}|z_{1:k+N})$ denoted by $F_{N+1}(\bar{u}_{N+1}, z_{k+1:k+N})$. Assembling the constants into K_{N+1}' shows that (7) also holds at $N + 1$. ■

REFERENCES

[1] H. G. Okuno and K. Nakadai, "Robot audition: Its rise and perspectives," in *2015 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, April 2015, pp. 5610–5614.

[2] S. Thrun, W. Burgard, and D. Fox, *Probabilistic Robotics*. The MIT Press, 2005.

[3] F. Bourgault, A. Makarenko, S. Williams, B. Grocholsky, and H. Durrant-Whyte, "Information based adaptive robotic exploration," in *IEEE/RSJ Int. Conf. on Intelligent Robots and Systems, (IROS)*, Lausanne, Switzerland, 2002.

[4] G. Bustamante, P. Danès, T. Forgue, A. Podlubne, and J. Manhès, "An information based feedback control for audio-motor binaural localization," *Autonomous Robots*, Jun 2017.

[5] C. Forster, M. Pizzoli, and D. Scaramuzza, "Appearance-based active, monocular, dense reconstruction for micro aerial vehicles," in *Proceedings of Robotics: Science and Systems*, Berkeley, USA, July 2014.

[6] S. Huang, N. M. Kwok, G. Dissanayake, Q. P. Ha, and G. Fang, "Multi-step look-ahead trajectory planning in slam: Possibility and necessity," in *Proceedings of the 2005 IEEE International Conference on Robotics and Automation*, April 2005, pp. 1091–1096.

[7] B. Deutsch, M. Zobel, J. Denzler, and H. Niemann, "Multi-step entropy based sensor control for visual object tracking," *Pattern Recognition*, pp. 359–366, 2004.

[8] M. Kumon, K. Fukushima, S. Kunimatsu, and M. Ishitobi, "Motion planning based on simultaneous perturbation stochastic approximation for mobile auditory robots," in *IEEE/RSJ Int. Conf. on Intelligent Robots and Systems (IROS)*, Taipei, Taiwan, 2010.

[9] C. Schymura, J. Diego, R. Grajales, and D. Kolossa, "Monte carlo exploration for active binaural localization," in *IEEE Int. Conf. on Acoustics, Speech, and Signal Processing (ICASSP)*, New Orleans, United States, Mar. 2017.

[10] E. Vincent, A. Sini, and F. Charpillet, "Audio source localization by optimal control of a mobile robot," in *IEEE Int. Conf. on Acoustics, Speech and Signal Processing (ICASSP)*, Brisbane, Australia, 2015.

[11] Q. V. Nguyen, F. Colas, E. Vincent, and C. F., "Long-term robot motion planning for active sound source localization with monte carlo tree search," in *Hands-free Speech Communication and Microphone Arrays (HSCMA)*, San Francisco, United States, Mar. 2017.

[12] A. Portello, P. Danès, S. Argentieri, and S. Pledel, "HRTF-based source azimuth estimation and activity detection from a binaural sensor," in *IEEE/RSJ Int. Conf. on Intelligent Robots and Systems (IROS)*, Tokyo, Japan, 2013.

[13] A. Portello, G. Bustamante, P. Danès, J. Piat, and J. Manhès, "Active localization of an intermittent sound source from a moving binaural sensor," in *Forum Acustium (FA)*, Krakow, Poland, 2014.

[14] N. Aaronson and W. Hartmann, "Testing, correcting, and extending the woodworth model for interaural time difference," *The Journal of the Acoustical Society of America*, vol. 135, pp. 817–823, 2014.

[15] T. Cover and J. Thomas, *Elements of Information Theory*. Wiley, 1991.

[16] S. Särkkä, J. Hartikainen, L. Svensson, and F. Sandblom, "Gaussian process quadratures in nonlinear sigma-point filtering and smoothing," in *17th International Conference on Information Fusion (FUSION)*, July 2014, pp. 1–8.

[17] M. Bartholomew-Biggs, S. Brown, B. Christianson, and L. Dixon, "Automatic differentiation of algorithms," *Journal of Computational and Applied Mathematics*, vol. 124, no. 1–2, pp. 171 – 190, 2000, numerical Analysis 2000. Vol. IV: Optimization and Nonlinear Equations.

[18] J. Manyika, "An information-theoretic approach to data fusion and sensor management," Ph.D. dissertation, University of Oxford, 1993.

[19] G. Bustamante, P. Danès, T. Forgue, and A. Podlubne, "Towards information-based feedback control for binaural active localization," in *IEEE Int. Conf. on Acoustics, Speech, and Signal Processing (ICASSP)*, Shanghai, China, 2016.