How Can Audio-Motor Binaural Localization Be Made "Active"?

Patrick Danès
LAAS-CNRS & Toulouse III - Paul Sabatier University ~ Toulouse ~ France

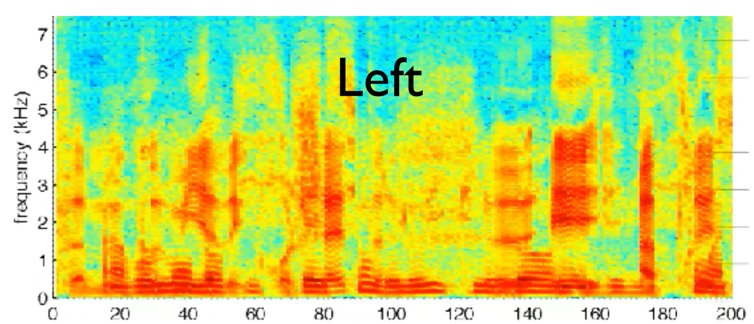
theory developed in collaboration with G. Bustamante (PhD2017) on top of A. Portello's work (PhD2013)
implementation thanks to involvement of T. Forgue & A. Podlubne (Two!Ears'oes) and J. Manhès

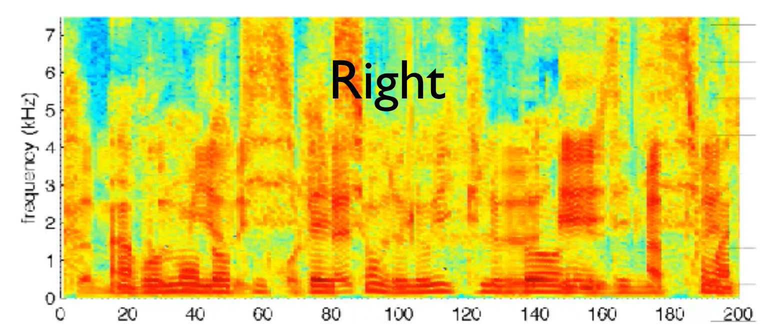
Context

Audio-Motor Localization of Source (E) from a Binaural Head (R1,R2)

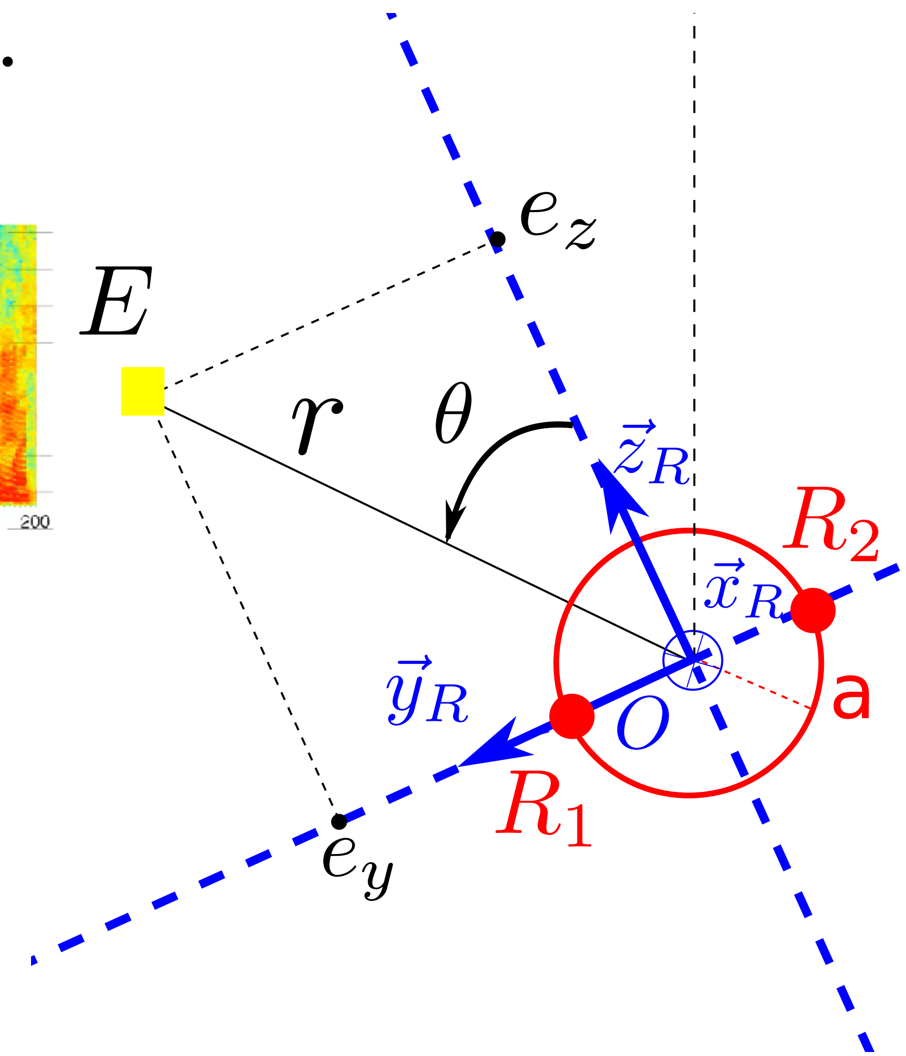
- Estimate state…
  - $x = (e_y, e_z)^T$: head-to-source position

- … by combining the sensed binaural signals…
  - z : left(R1) and right(R2) spectrograms

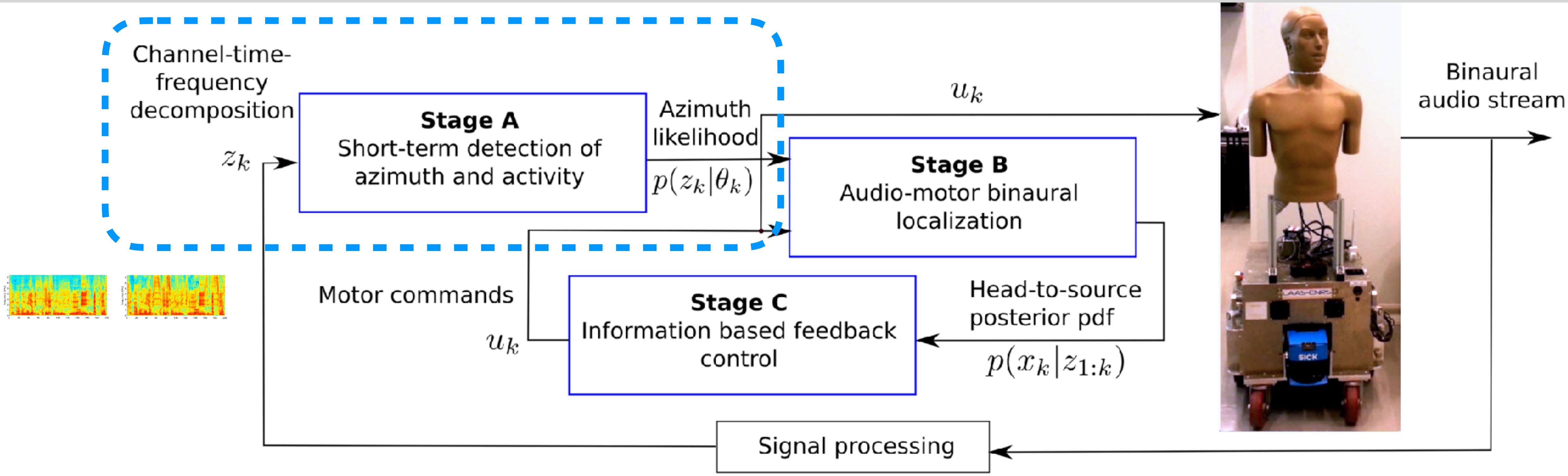




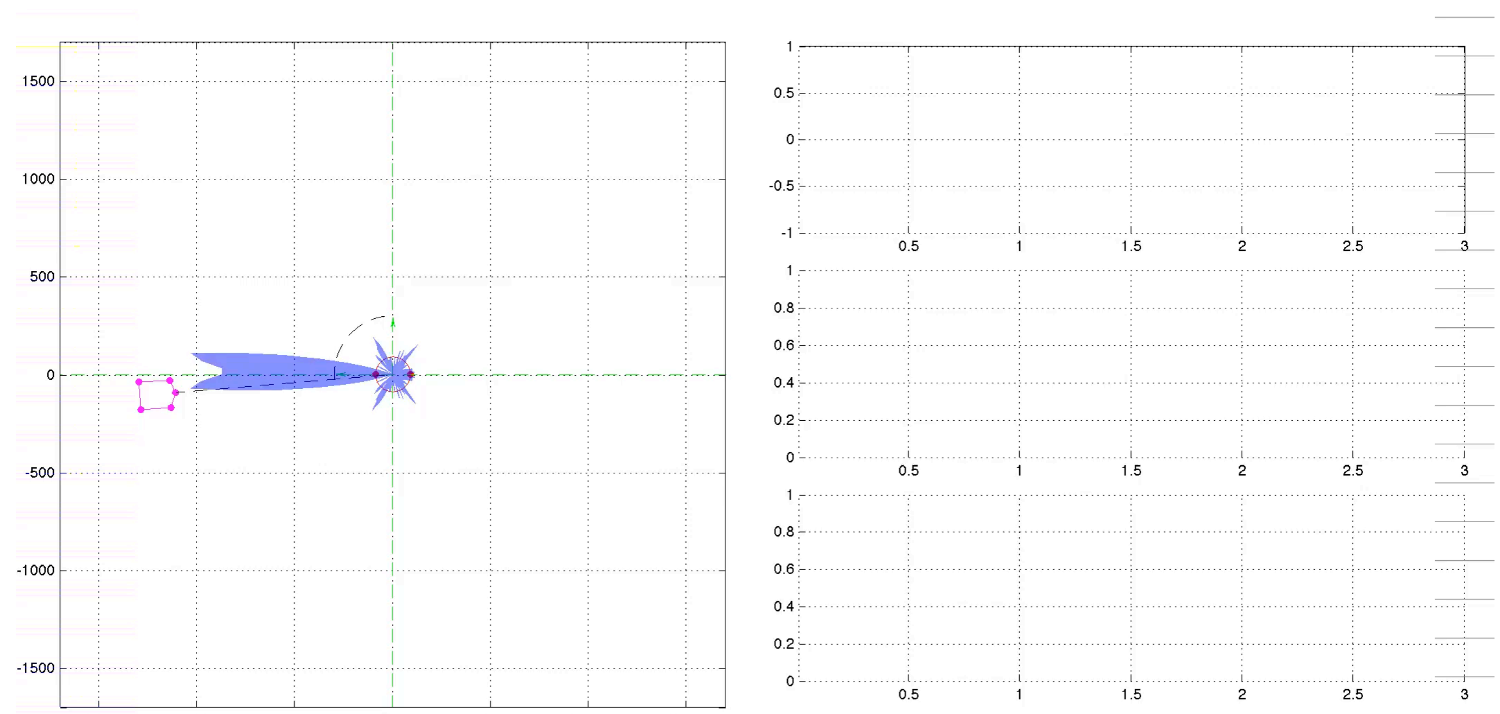
- … with the motor commands of the sensor
  - $u = (T_y, T_z, \phi)$: head translations and rotation



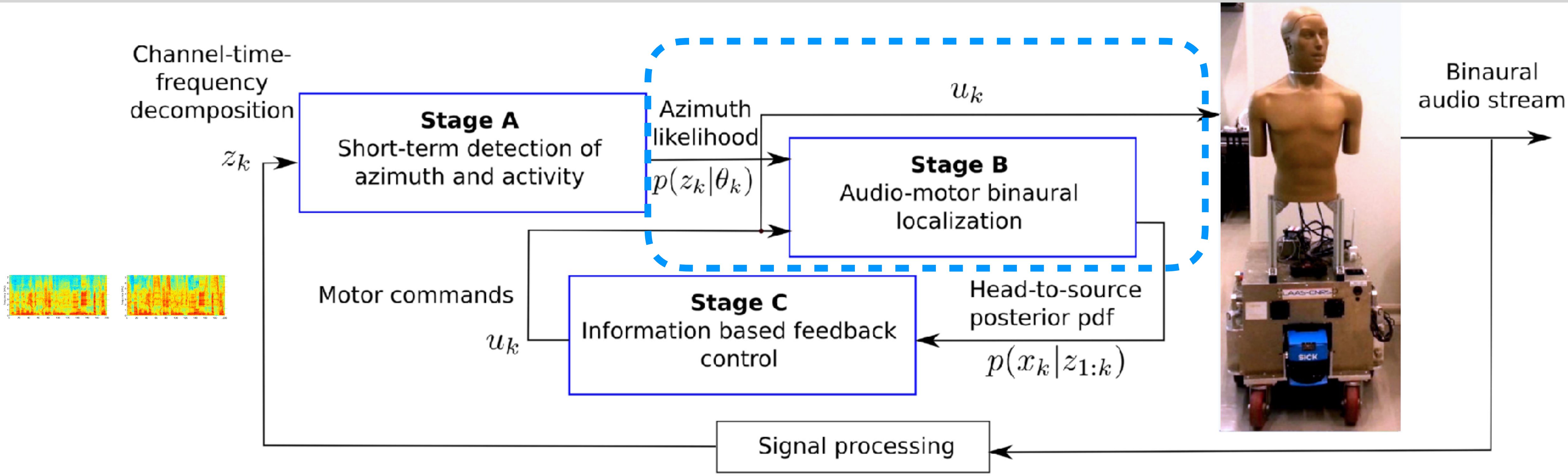
A Three-Layer Framework to Binaural Active Localization



(A) HRTF-based ML Source Azimuth Estimation over small time snippets



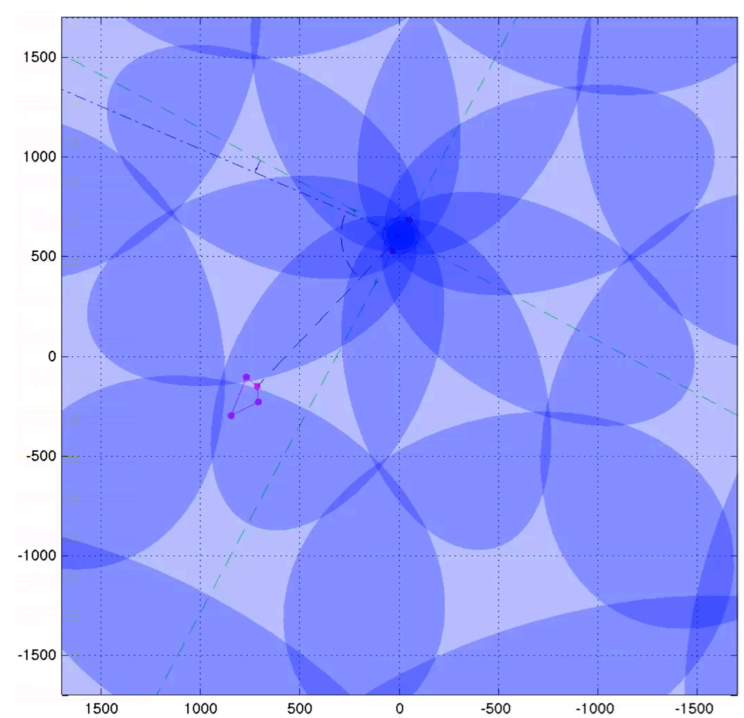
A Three-Layer Framework to Binaural Active Localization



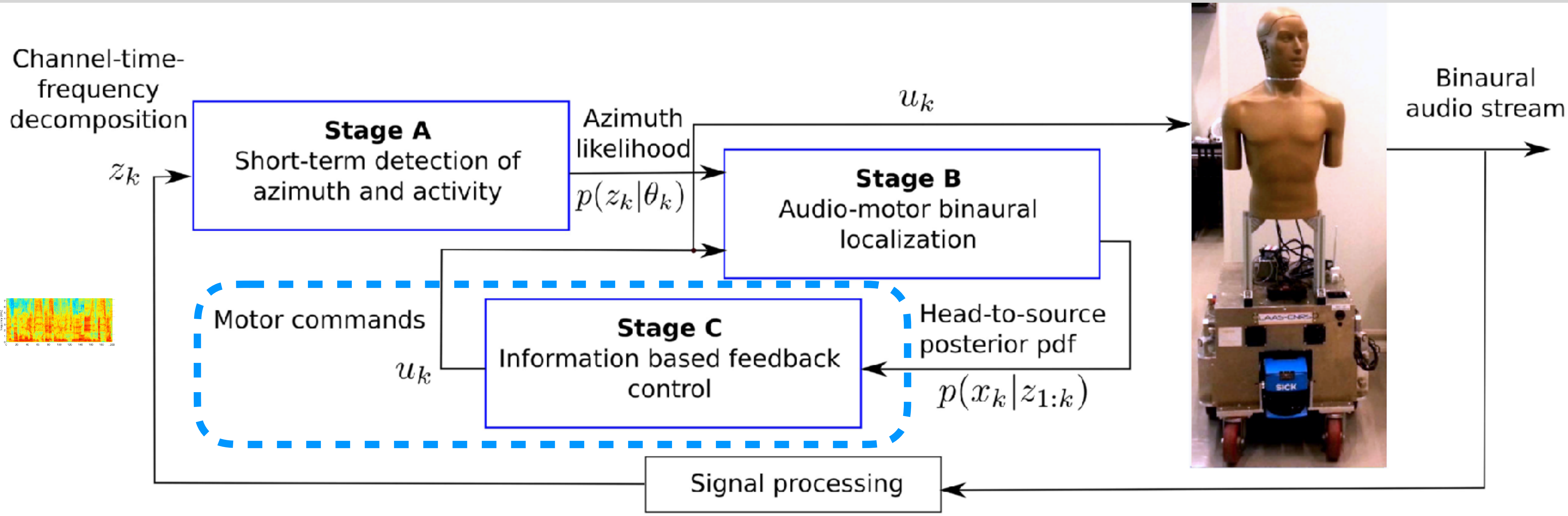
(B) Gaussian sum UKF Audio-Motor Localization

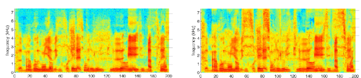
- front-back disambiguation
- range recovery

$$p(x_k|z_{1:k}) \approx \sum_{j=1}^{J_k} \gamma_k^j \mathcal{N}(x_k; \hat{x}_{k|k}^j, P_{k|k}^j)$$

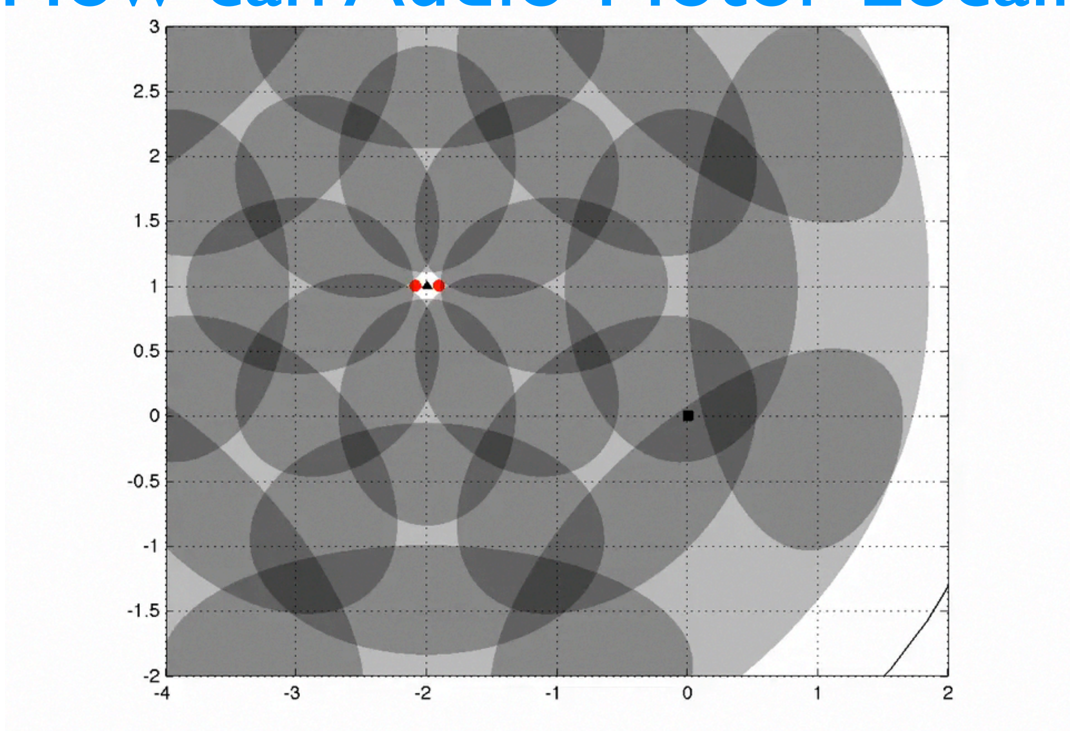


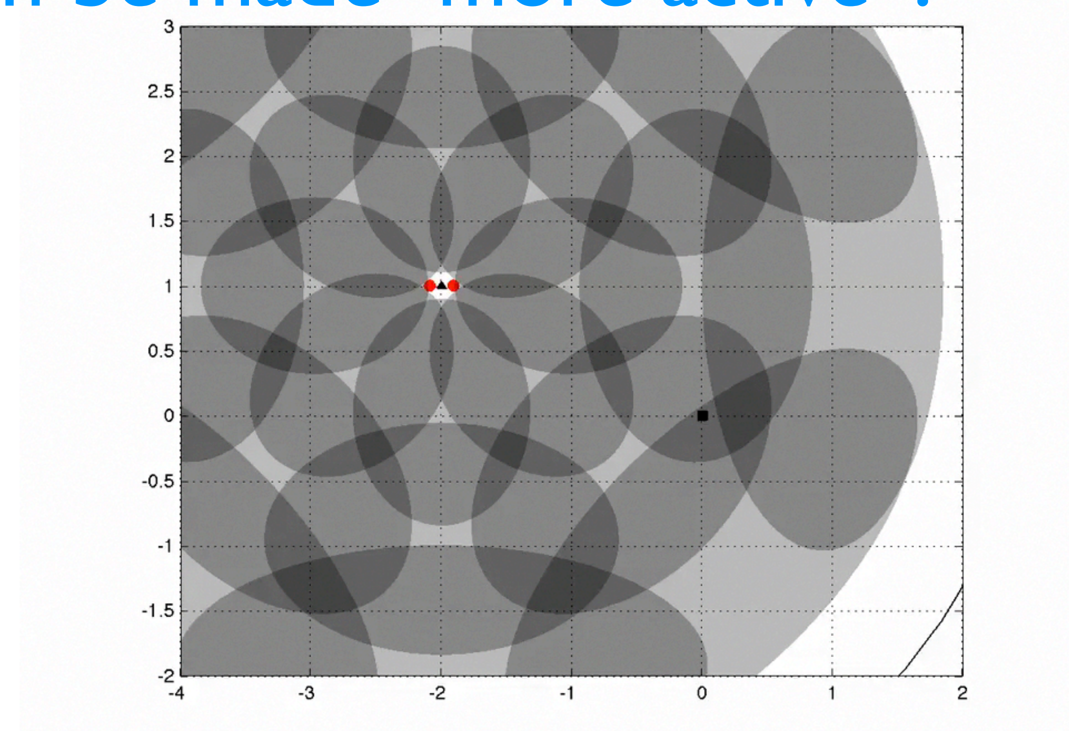
A Three-Layer Framework to Binaural Active Localization





(C) How can Audio-Motor Localization be made "more active"?





Problem Statement

Multi-step-ahead information-based feedback control

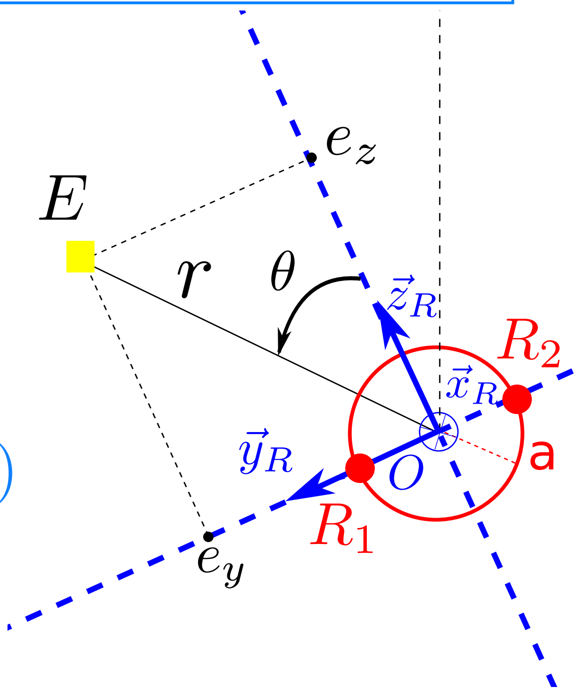
Given $p(x_k|z_{1:k}) \approx \sum_{j=1}^{J_k} \gamma_k^j \mathcal{N}(x_k; \hat{x}_{k|k}^j, P_{k|k}^j)$, find $\bar{u}_N^* = u_k^\star : u_{k+N-1}^*$ (to be applied between k and $k+N$) such that $\mathbb{E}_{z_{k+1:k+N}|z_{1:k}} h(x_{k+N}|z_{1:k+N})$ is minimum, with $h(x_{k+N}|z_{1:k+N})$ the entropy of $p(x_{k+N}|z_{1:k+N})$

Multi-step-ahead information-based feedback control

Given $p(x_k|z_{1:k}) \approx \sum_{j=1}^{J_k} \gamma_k^j \mathcal{N}(x_k; \hat{x}_{k|k}^j, P_{k|k}^j)$, find $\bar{u}_N^* = u_k^\star : u_{k+N-1}^*$ (to be applied between k and $k+N$) such that $\mathbb{E}_{z_{k+1:k+N}|z_{1:k}} h(x_{k+N}|z_{1:k+N})$ is minimum, with $h(x_{k+N}|z_{1:k+N})$ the entropy of $p(x_{k+N}|z_{1:k+N})$

Simplifying assumptions

- $p(x_k|z_{1:k}) \approx \mathcal{N}(x_k; \hat{x}_{k|k}, P_{k|k})$
- $z_k = \bar{l}(\theta_k) + v_k,\ \theta_k = -\text{atan2}(e_y, e_z),\ v_k \sim \mathcal{N}(0, R_k)$ with R_k independent of hidden (e_{y_k}, e_{z_k})
  - Woodworth-Scholsberg (farfield) approximation of ITD

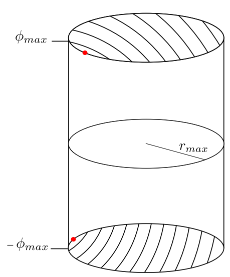


Outline

- One-step-ahead control: theory, qualitative insights, experiments
- N-step-ahead control: theory, simulated experiments

In view of the UKF equations

- $p(x_{k+1}|z_{1:k+1}) \approx \mathcal{N}\big(x_{k+1}; \hat{x}_{k+1|k+1}, P_{k+1|k+1}\big)$, with $P_{k+1|k+1}$ independent of z_{k+1}
- $h(x_{k+1}|z_{1:k+1}) = a \log\det P_{k+1|k+1} + b \,(\text{with } a > 0) \;= \mathbb{E}_{z_{k+1}|z_{1:k}} h(x_{k+1}|z_{1:k+1})$
- $(T_y^*, T_z^*, \phi^*) = \arg \min_{(T_y,T_z,\phi)\in\mathcal{T}\times\mathcal{R}} h(x_{k+1}|z_{1:k+1})$



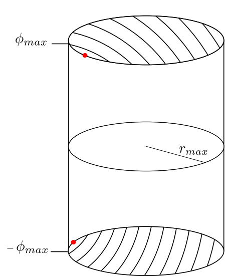
$$= \arg \max_{(T_y,T_z,\phi)\in\mathcal{T}\times\mathcal{R}} F_1(T_y,T_z,\phi), \text{ with } F_1(T_y,T_z,\phi) = h(z_{k+1}|z_{1:k})$$

and $h(z_{k+1}|z_{1:k})$ entropy of $p(z_{k+1}|z_{1:k})$

One-step-ahead Solution Basics [Autonomous Robots 2018]

In view of the UKF equations

- $p(x_{k+1}|z_{1:k+1}) \approx \mathcal{N}\big(x_{k+1}; \hat{x}_{k+1|k+1}, P_{k+1|k+1}\big)$, with $P_{k+1|k+1}$ independent of z_{k+1}
- $h(x_{k+1}|z_{1:k+1}) = a \log\det P_{k+1|k+1} + b\,(\text{with } a>0) = \mathbb{E}_{z_{k+1}|z_{1:k}} h(x_{k+1}|z_{1:k+1})$
- $(T_y^*, T_z^*, \phi^*) = \arg\min_{(T_y,T_z,\phi)\in\mathcal{T}\times\mathcal{R}} h(x_{k+1}|z_{1:k+1})$



$$= \arg\max_{(T_y,T_z,\phi)\in\mathcal{T}\times\mathcal{R}} F_1(T_y,T_z,\phi), \text{ with } F_1(T_y,T_z,\phi) = h(z_{k+1}|z_{1:k})$$

and $h(z_{k+1}|z_{1:k})$ entropy of $p(z_{k+1}|z_{1:k})$

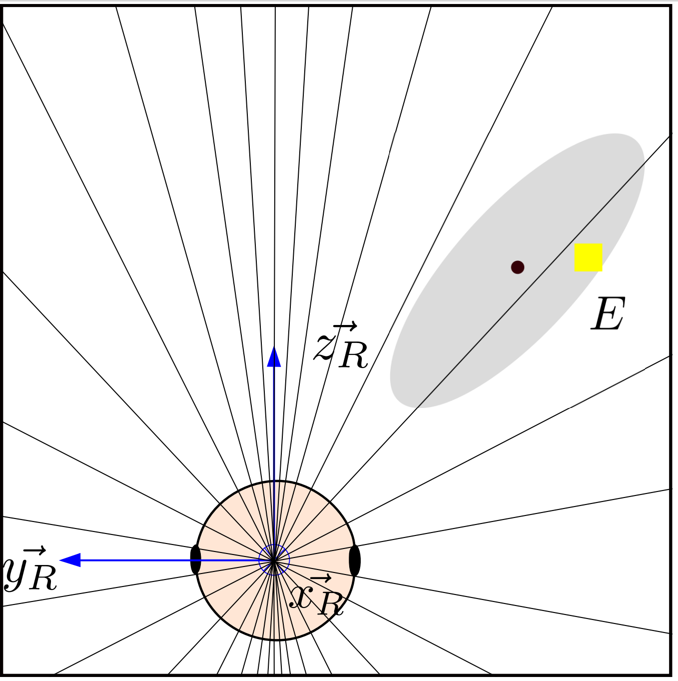
Facts

- $p(z_{k+1}|z_{1:k}) \approx \mathcal{N}\big(z_{k+1}; \hat{z}_{k+1|k}, S_{k+1|k}\big)$, so that

$$\bar{u}_1^\star = (T_y^*, T_z^*, \phi^*) = \arg\max_{((T_y,T_z),\phi)\in\mathcal{T}\times\mathcal{R}} F_1(T_y,T_z,\phi) = \log\det S_{k+1|k}$$

- $F_1(T_y,T_z,\phi)$ has no closed form, but its gradient can be approximated through Taylor expansions and the unscented transform

➟ The problem can be solved by the projected gradient algorithm

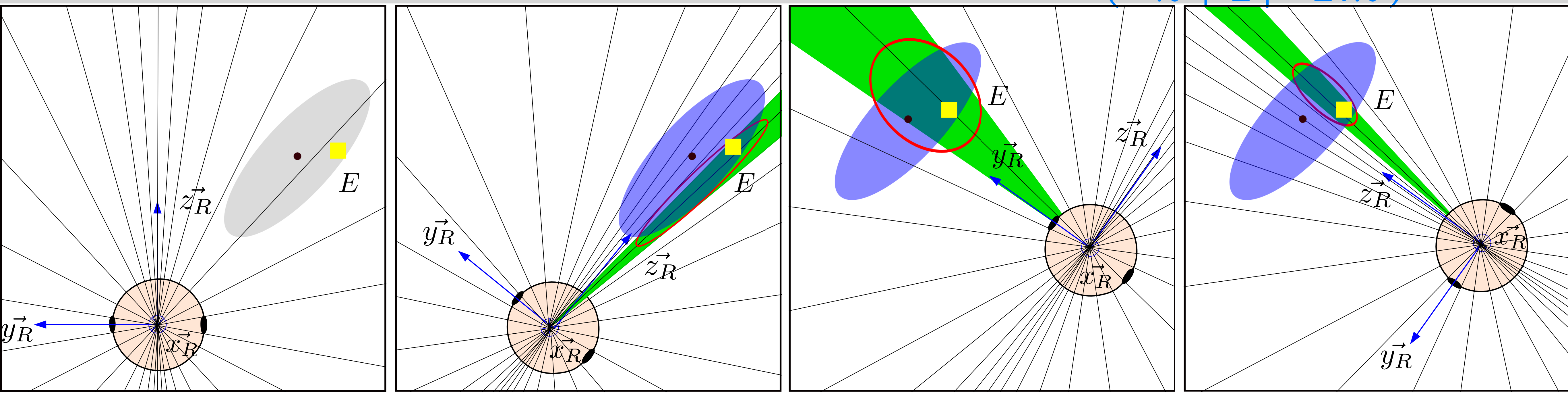


(a)

$\mathcal{F}:(O,\vec{e_x},\vec{e_y},\vec{e_z})$: binaural head. Measurement space: Woodworth iso-ITDs.

Sound source real position.

Confidence ellipse: current posterior head-to-source pdf.



(a) (b) (c) (d)

$\mathcal{F} : (O, \vec{e_x}, \vec{e_y}, \vec{e_z})$: binaural head. Measurement space: Woodworth iso-ITDs.

Sound source real position.

Confidence ellipse: current posterior head-to-source pdf.

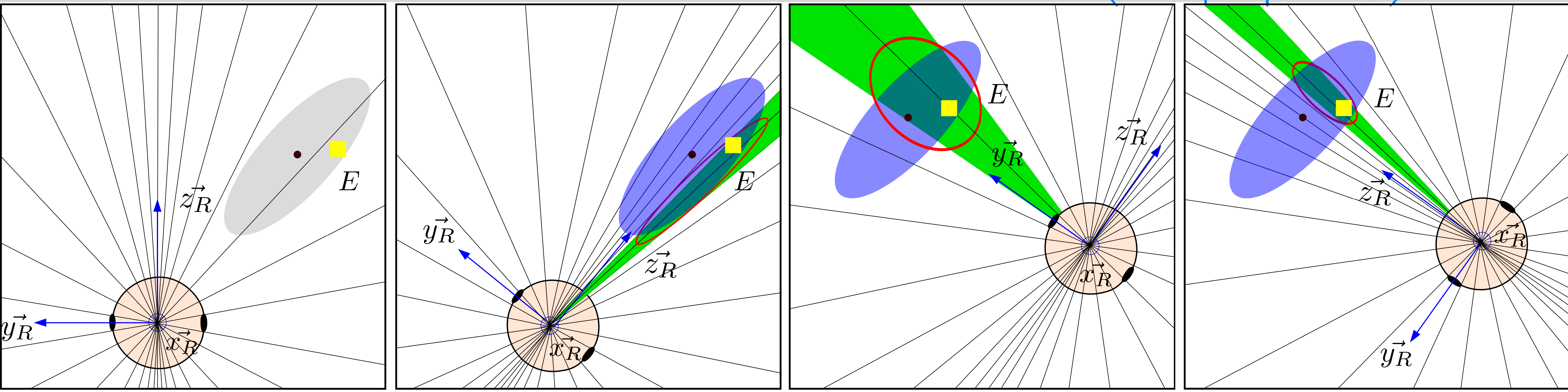
Next predicted state pdf, after applying T_y, T_z, ϕ to the head (with no dyn. noise).

Spatial uncertainty sector due to measurement noise.

Confidence ellipsoid associated to the next filtered state pdf (after incorporating the Woodworth ITD for the source position).

One-step-ahead Solution
Why and How to Maximize $h(z_{k+1}|z_{1:k})$?

Two!Ears

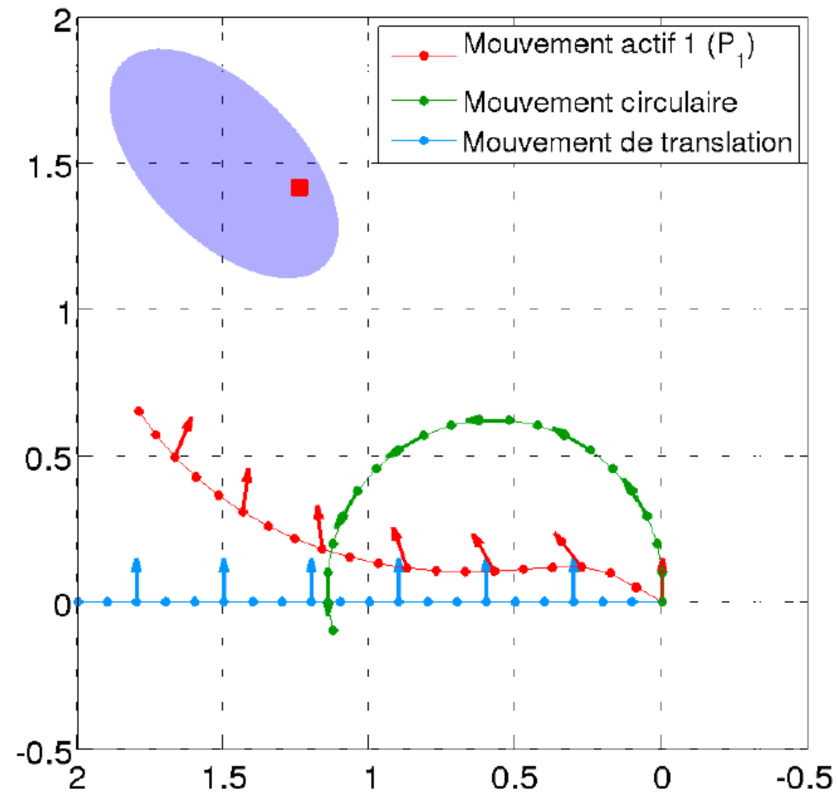


(a) (b) (c) (d)

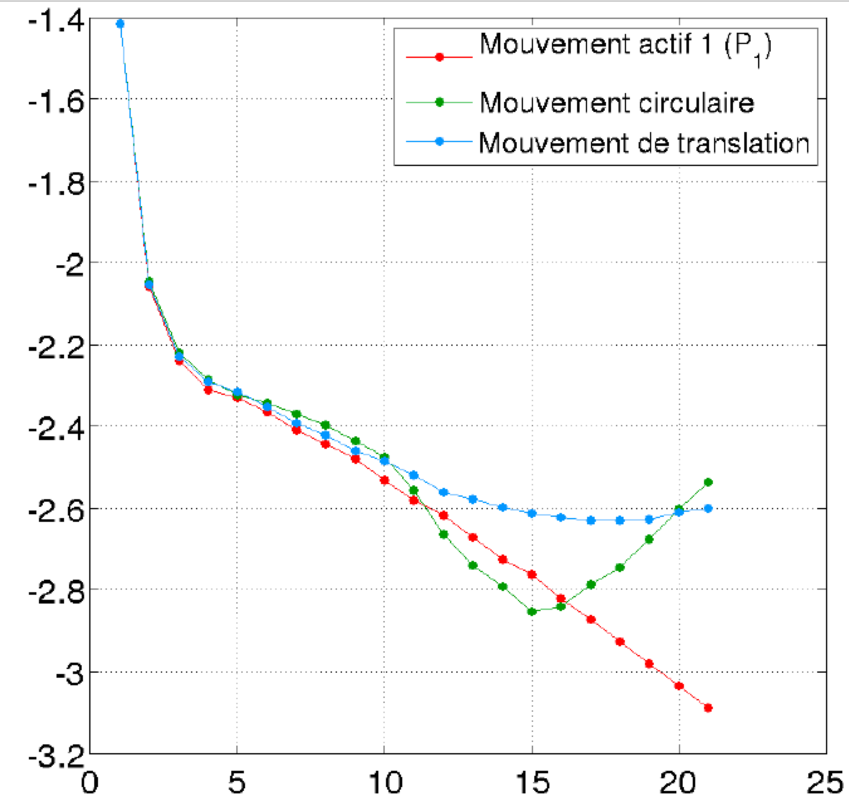
Some heuristical rules of thumb

- \- Do not orient the interaural axis towards the confidence ellipsoid
- \+ Orient the auditory fovea towards the confidence ellipsoid
- \- Drive the head center on the (line supported by) the major axis of the confidence ellipsoid
- \+ Drive the head center on the minor axis of the confidence ellipsoid
- \+ Get closer to the confidence ellipsoid
- ➡ so that the ellipsoid intersects as many iso-ITDs as possible

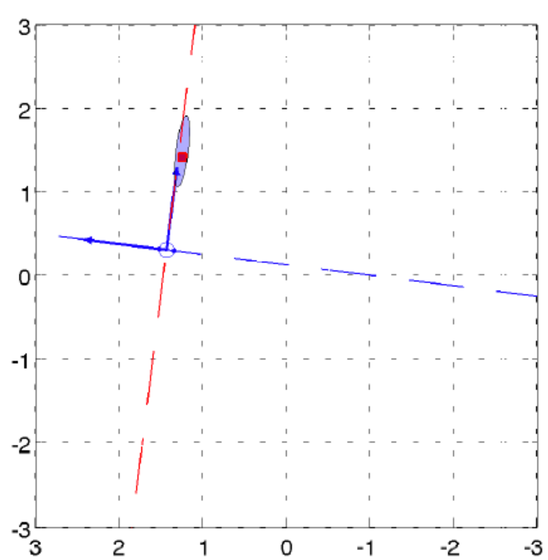
One-step-ahead Solution
Simulated Experiments



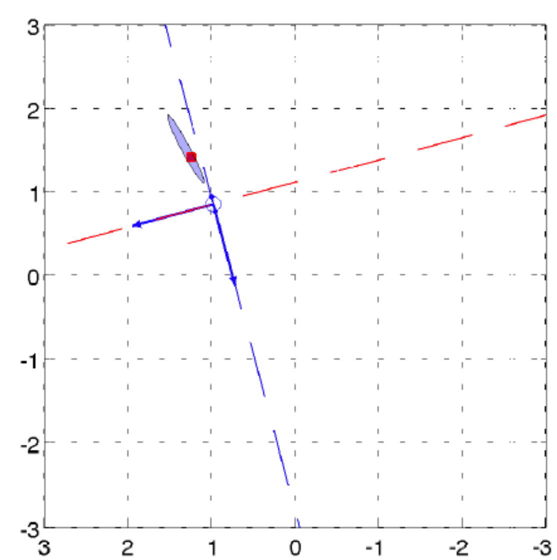
(a) Trajectoires dans le repère monde



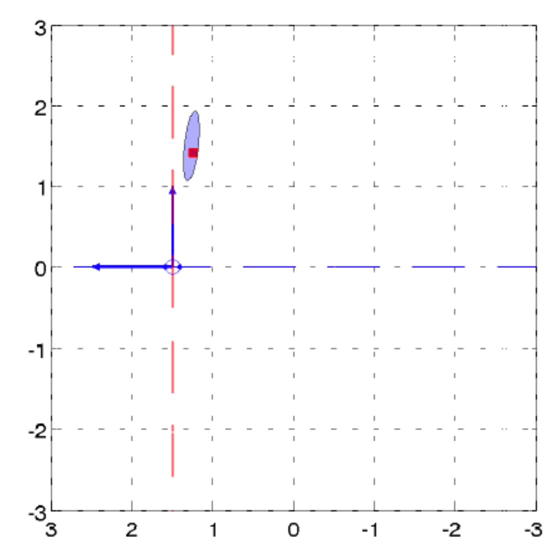
(b) Entropies



(c) Mouvement actif $k = 16$

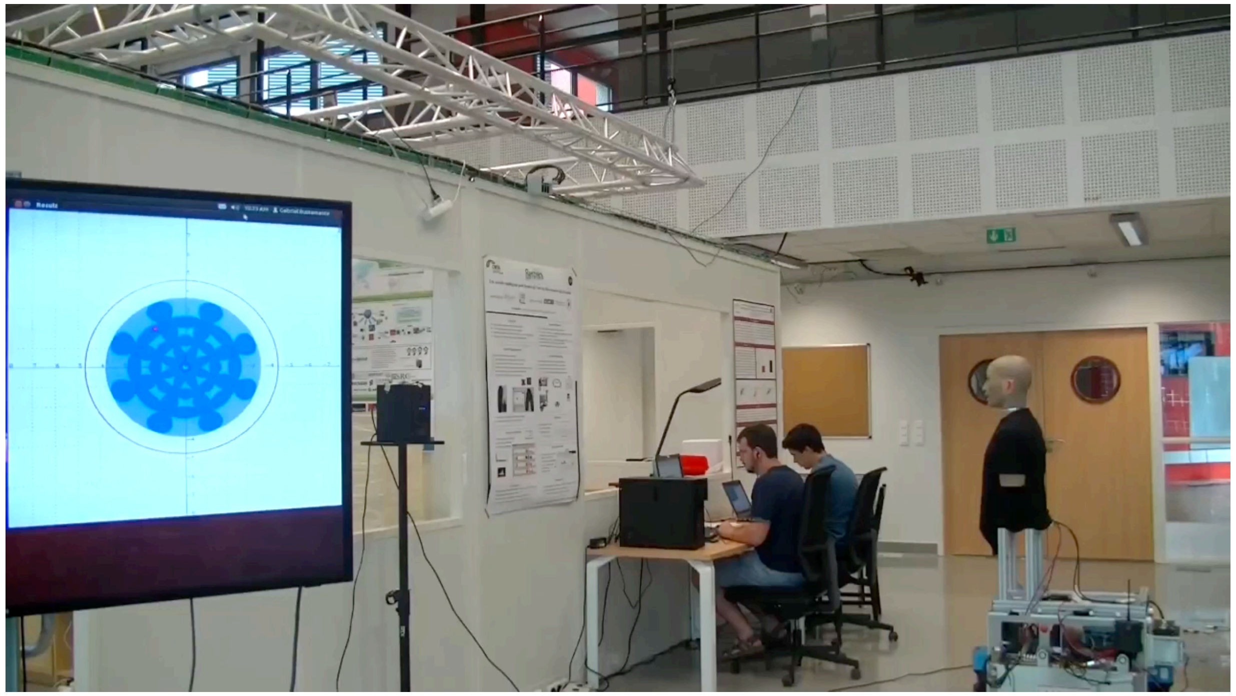


(d) Mouvement circulaire $k = 16$



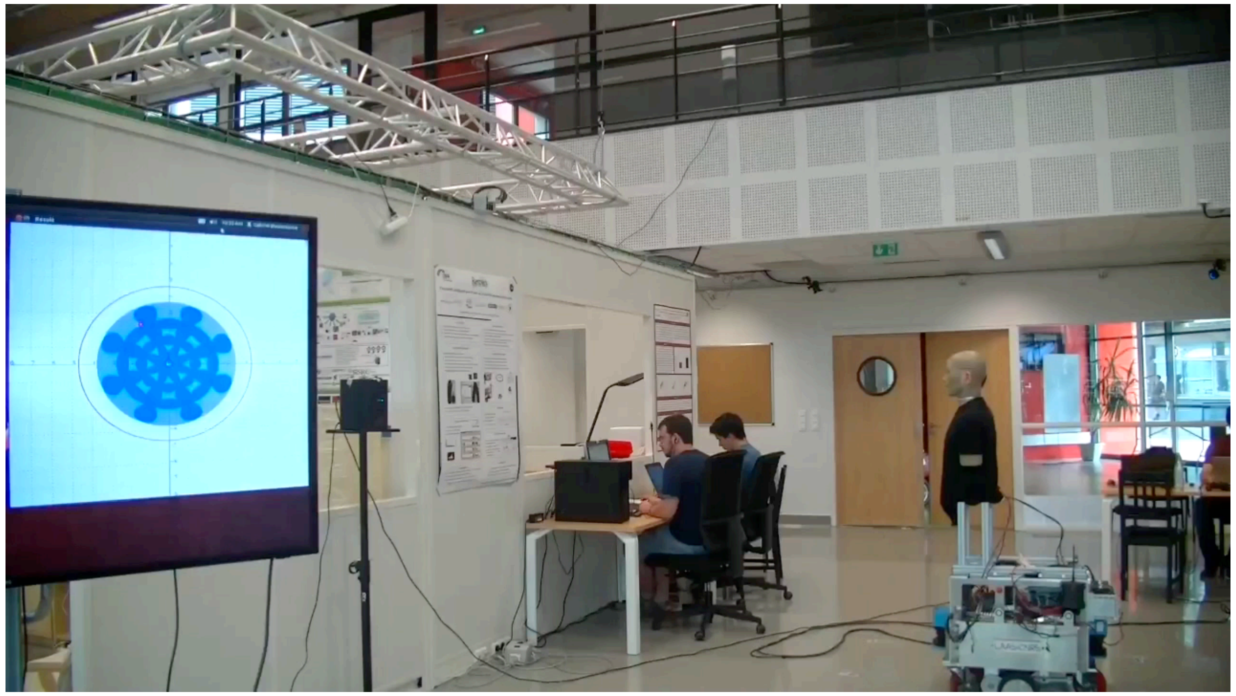
(e) Mouvement de translation $k = 16$

One-step-ahead Solution
Live Experiments



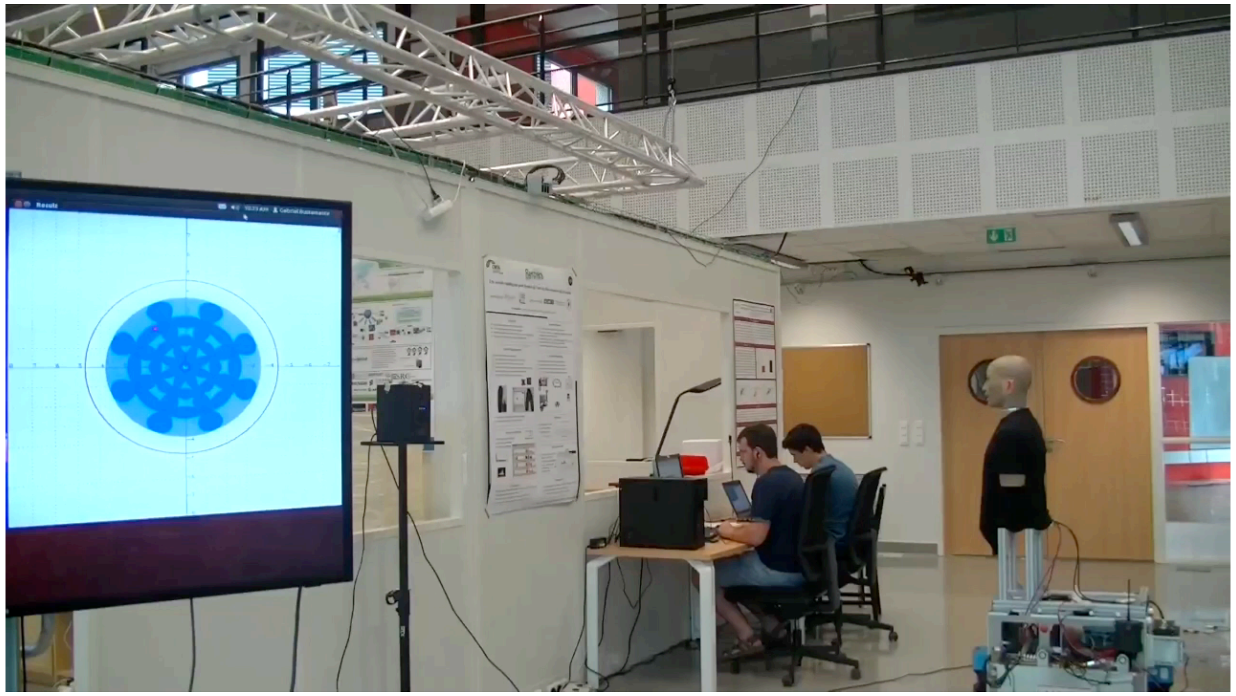
Open-loop head translation

Open-loop head rotation

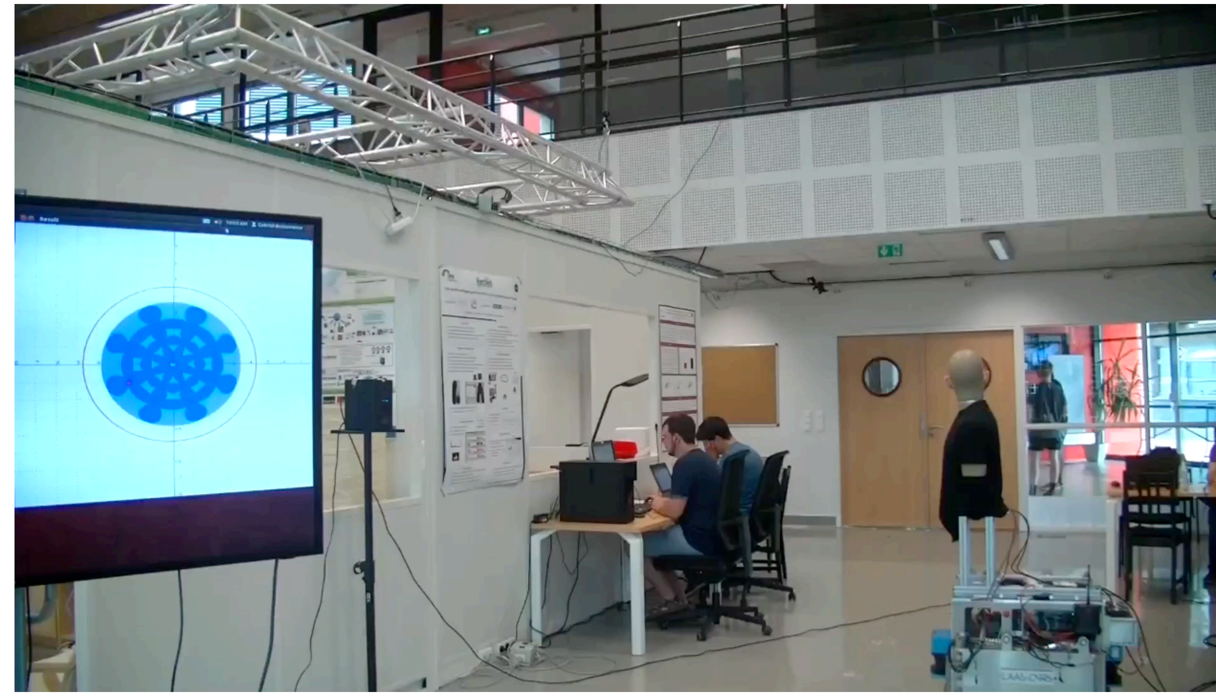


Open-loop circular nonholonomic motion

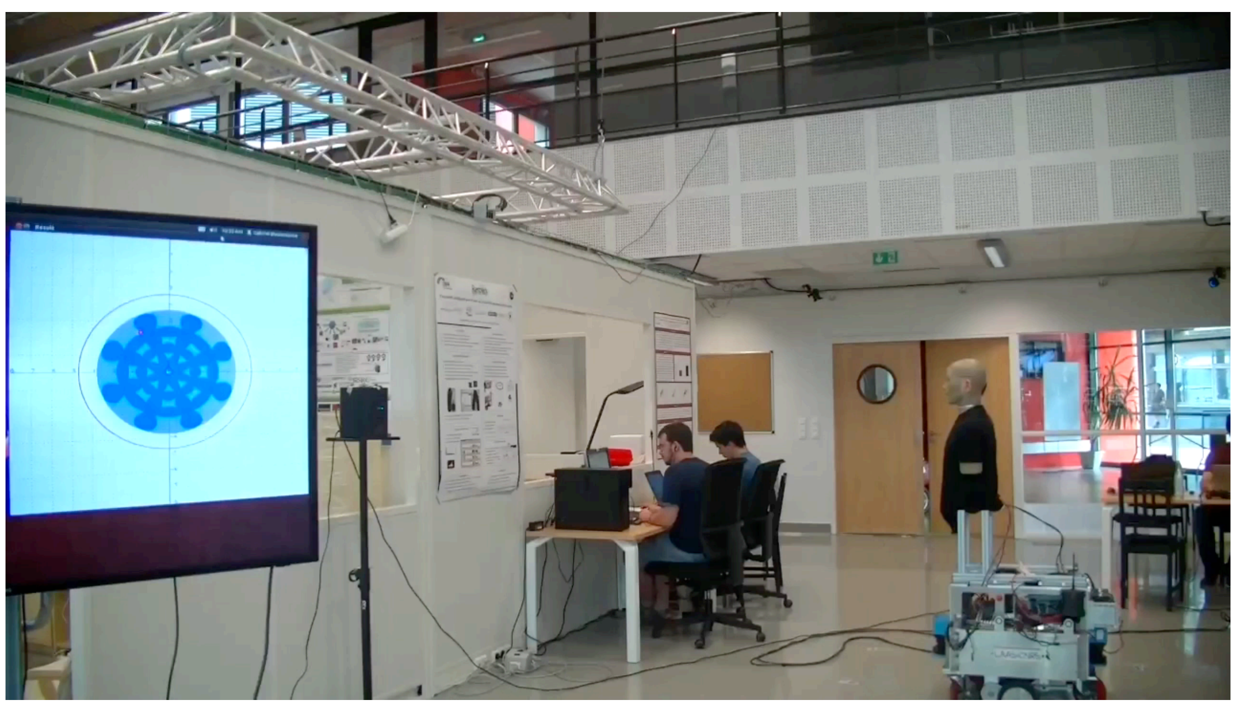
One-step-ahead Solution
Live Experiments



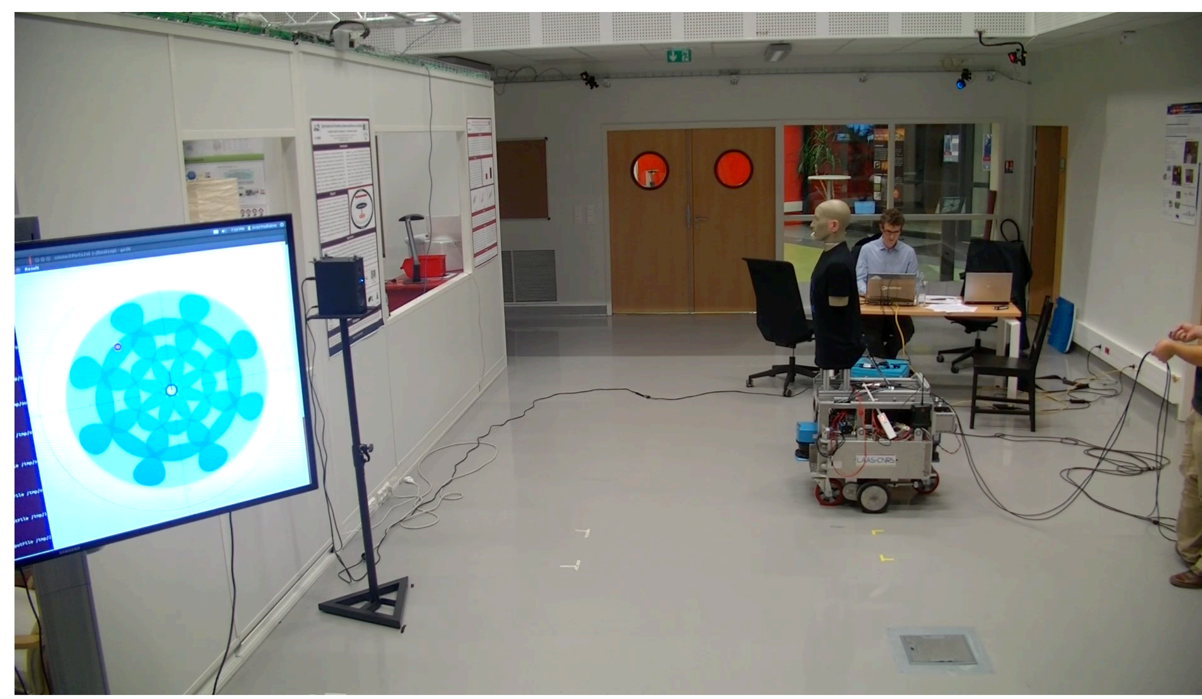
Open-loop head translation



Open-loop head rotation

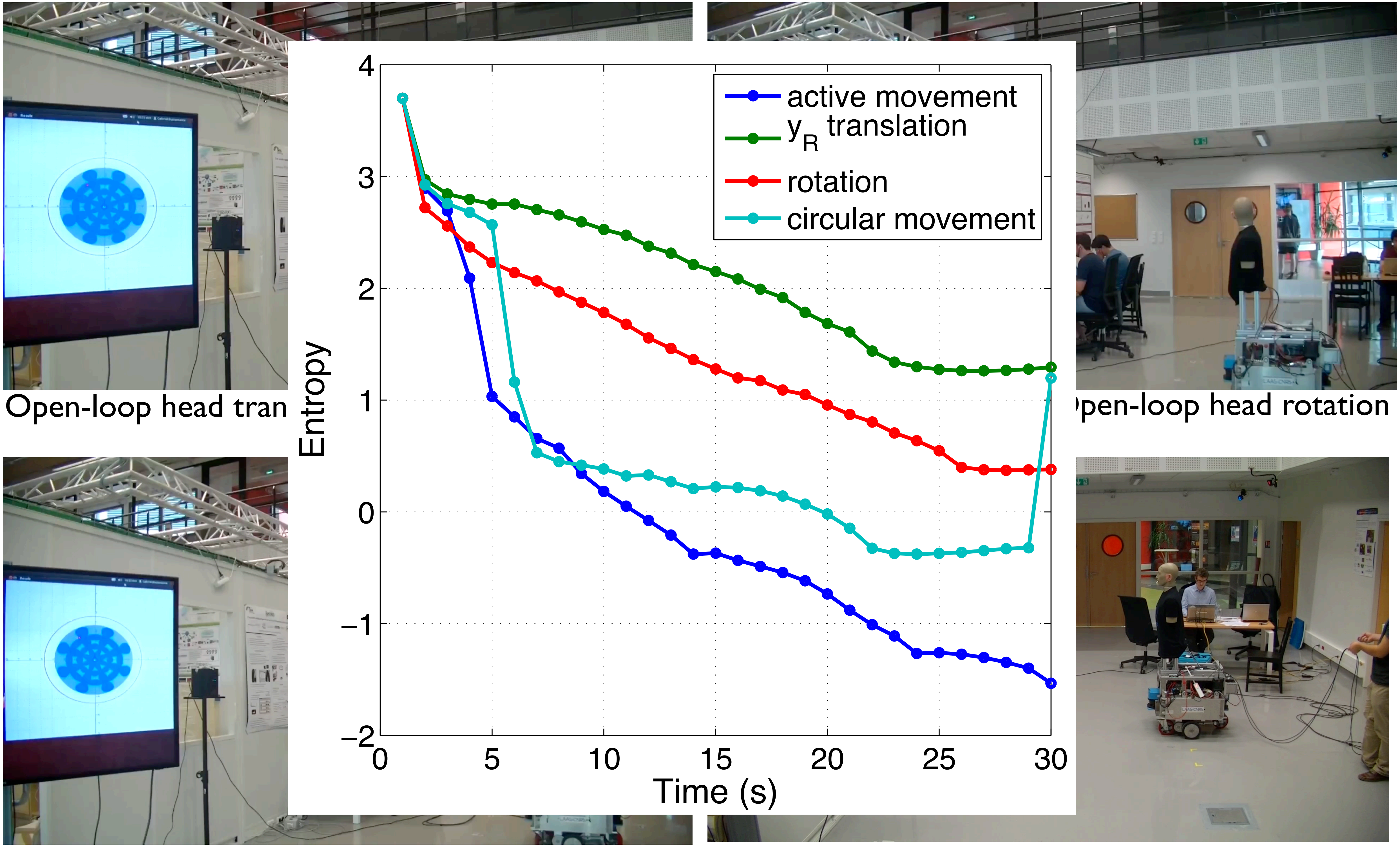


Open-loop circular nonholonomic motion



One-step-ahead optimal control

One-step-ahead Solution
Live Experiments



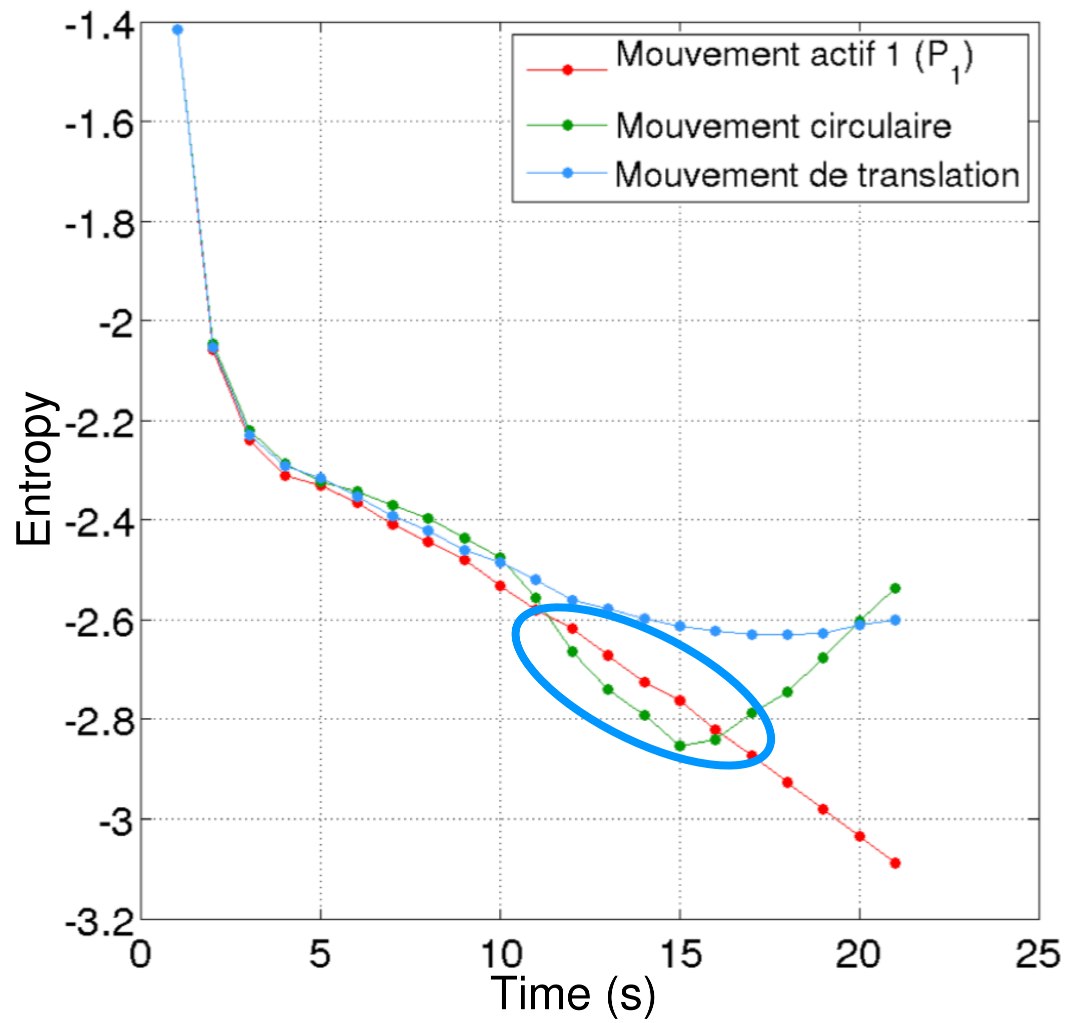
Open-loop head tran

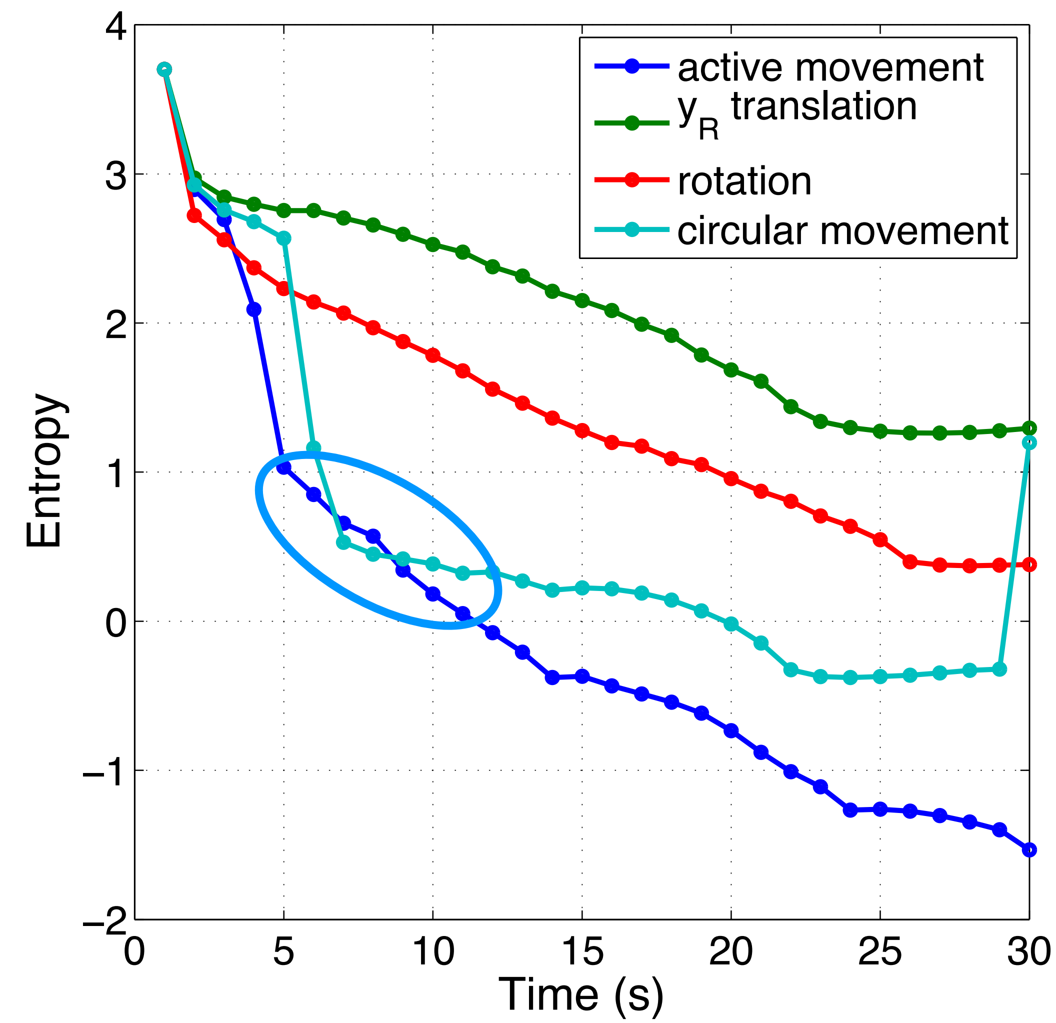
Open-loop head rotation

Open-loop circular nonholonomic motion

One-step-ahead optimal control

One-step-ahead problem is solved… … so, what may be wrong?





The sequence of N locally optimum one-step-ahead strategies may not be optimal ➟ let's investigate a N-step-ahead strategy!

N-step-ahead Solution Basics [IROS'2017]

Given $p(x_k|z_{1:k}) \approx \mathcal{N}(x_k; \hat{x}_{k|k}, P_{k|k})$, find the value of $\bar{u}_N = u_k : u_{k+N-1}$ which minimizes $J_N(\bar{u}_N) = \mathbb{E}_{z_{k+1:k+N}|z_{1:k}} h(x_{k+N}|z_{1:k+N})$

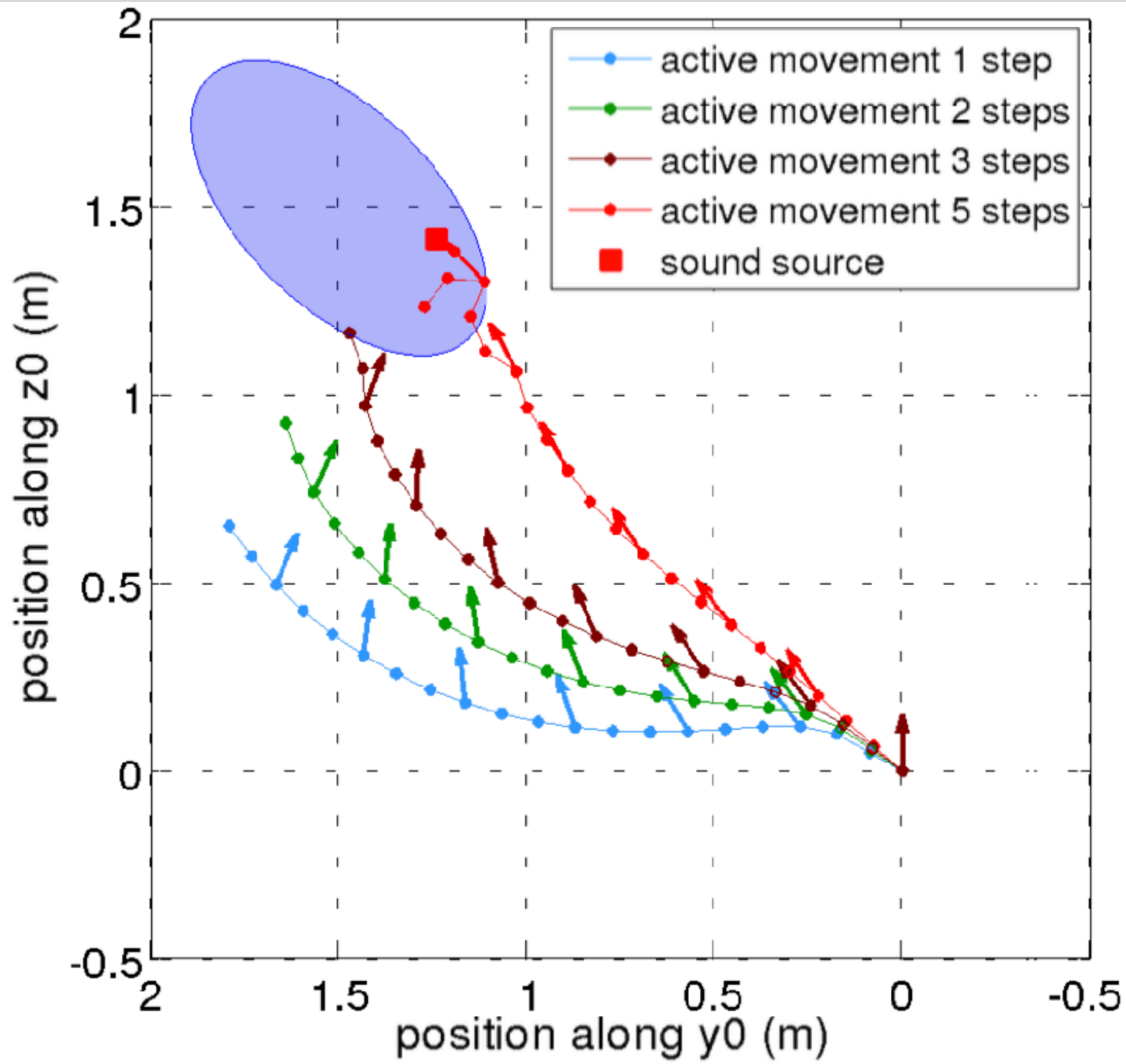
N-step-ahead Solution Basics [IROS'2017]

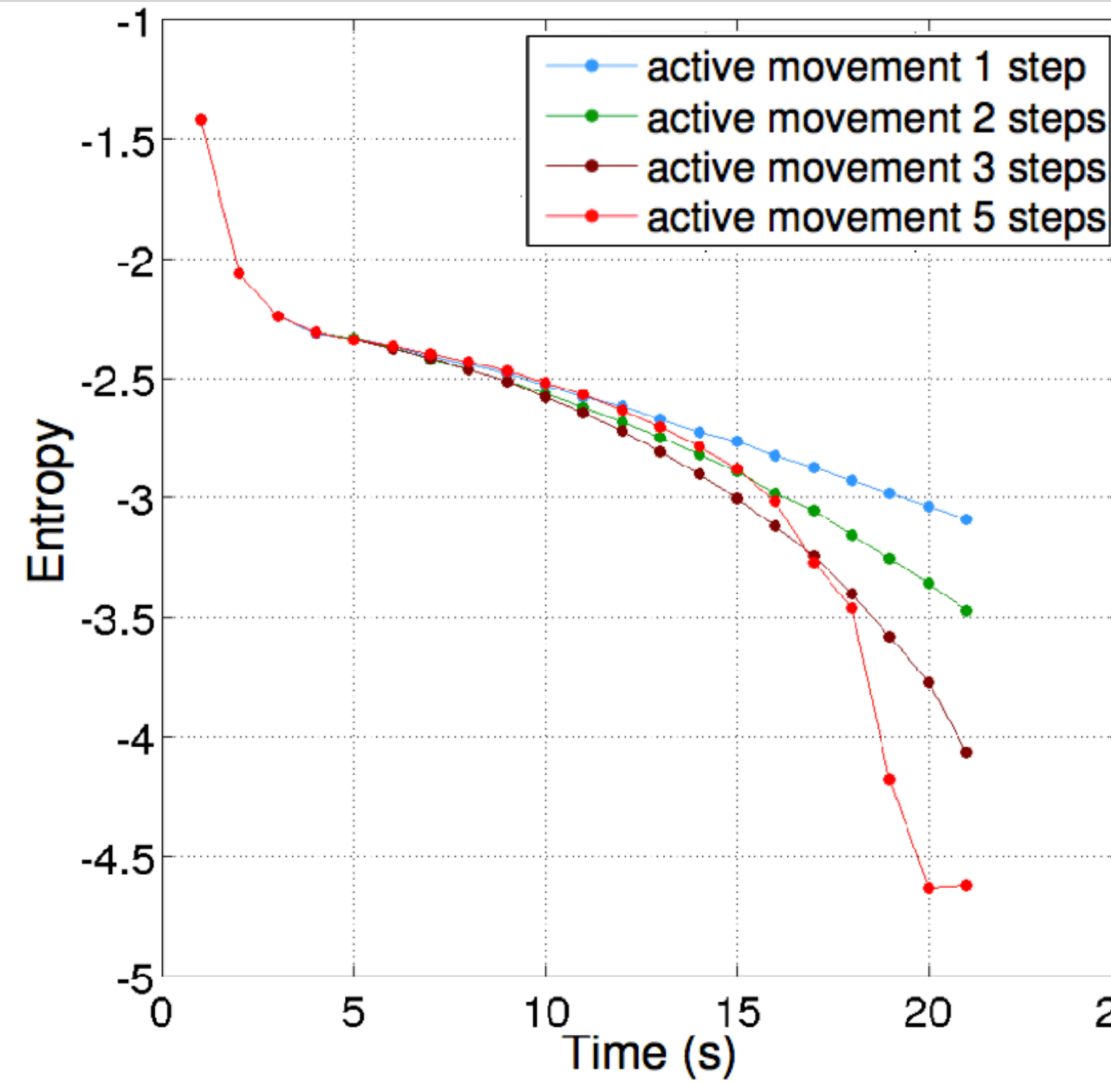
Given $p(x_k|z_{1:k}) \approx \mathcal{N}(x_k; \hat{x}_{k|k}, P_{k|k})$, find the value of $\bar{u}_N = u_k : u_{k+N-1}$ which minimizes $J_N(\bar{u}_N) = \mathbb{E}_{z_{k+1:k+N}|z_{1:k}} h(x_{k+N}|z_{1:k+N})$

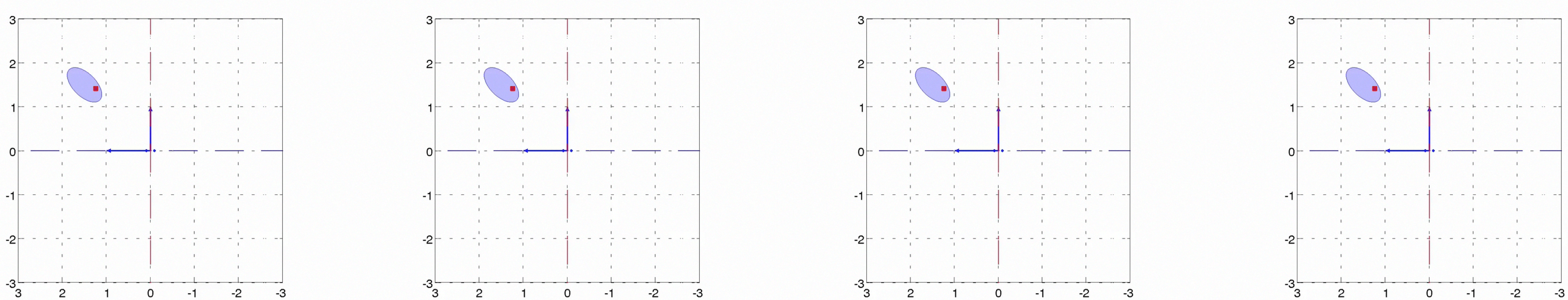
Some important (and muuuuch more tricky) facts

- $$J_N = K'_N - \underbrace{h(z_{k+1}|z_{1:k})}_{\triangleq F_1(\bar{u}_1)} - \sum_{i=2}^{N} \mathbb{E}_{z_{k+1:k+i-1}|z_{1:k}} \underbrace{\Big[h(z_{k+i}|z_{1:k+i-1})\Big]}_{\triangleq F_i(\bar{u}_i, z_{k+1:k+i-1})}$$
- By the Unscented Transform,
$$\mathbb{E}_{z_{k+1:k+i-1}|z_{1:k}} \Big[F_i(\bar{u}_i, z_{k+1:k+i-1})\Big] \approx \sum_{j=1}^{2(i-1)+1} W_j F_i(\bar{u}_i, Z_j)$$
- Finally,
$$\bar{u}_N^\star = \arg \min_{\bar{u}_N \in (\mathcal{T}\times\mathcal{R})^N} J_N(\bar{u}_N) \approx \arg \max_{\bar{u}_N \in (\mathcal{T}\times\mathcal{R})^N} F_1(\bar{u}_1) + \sum_{i=2}^{N} \sum_{j=1}^{2(i-1)+1} W_j F_i(\bar{u}_i, Z_j)$$
- Solution by means of the projected gradient algorithm
  - automatic differentiation: dual numbers algebra $v + \varepsilon d$, $\varepsilon^2 = 0$

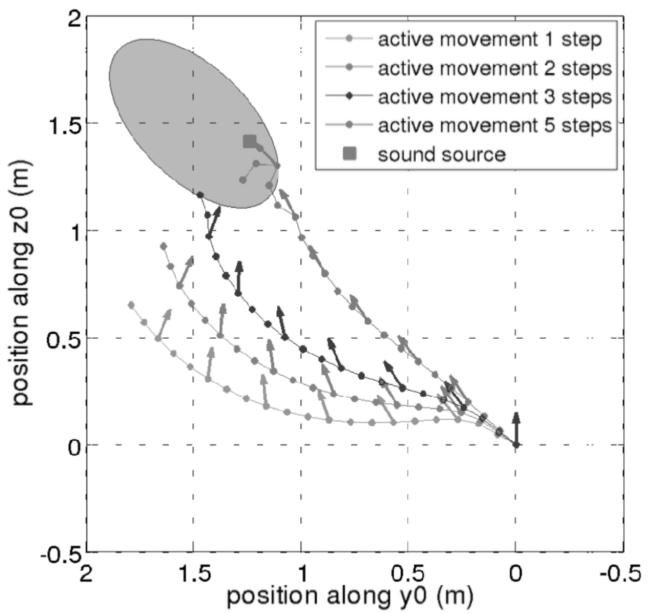
N-step-ahead Solution
Simulated Experiments

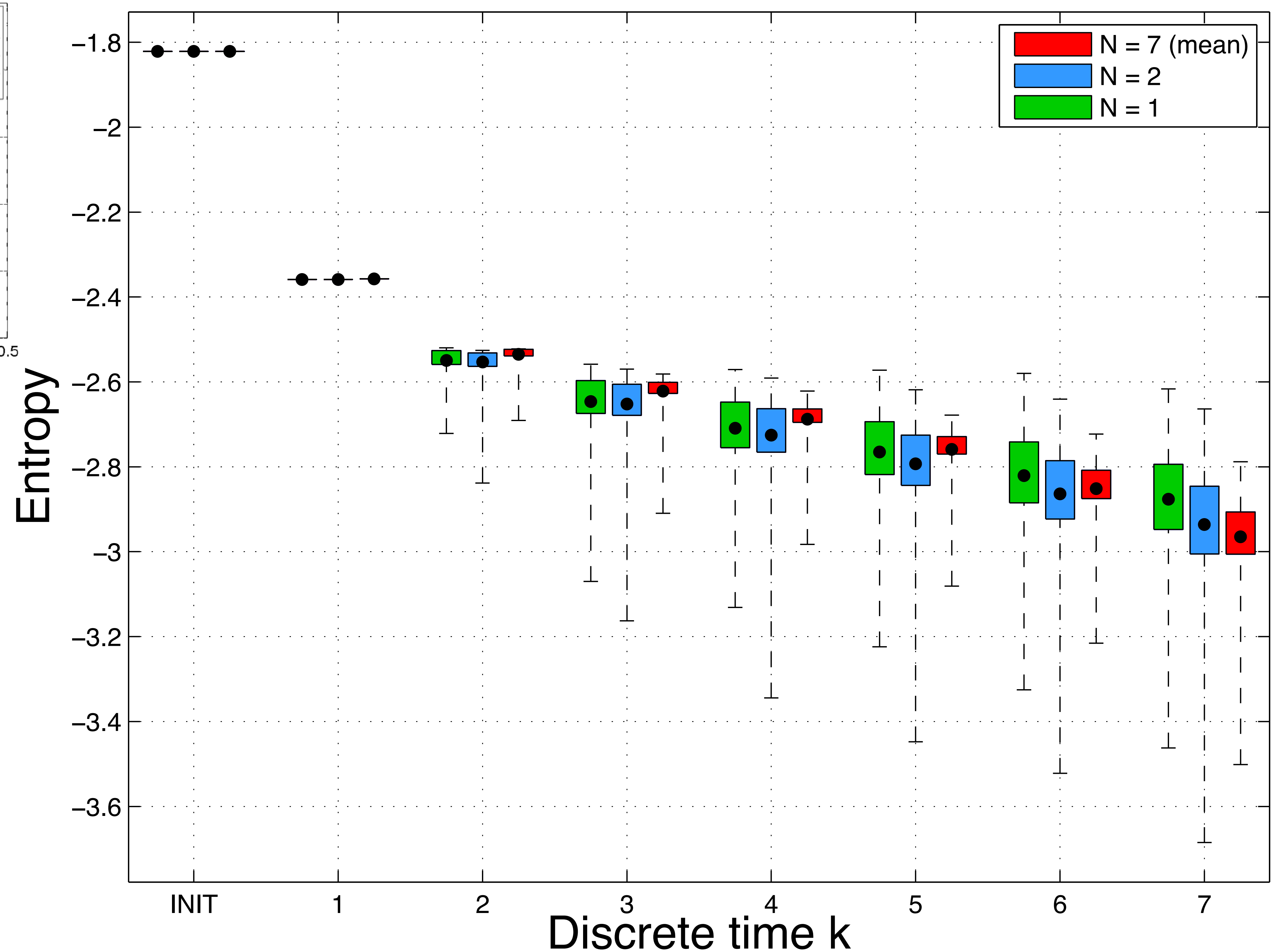






N-step-ahead Solution
Monte Carlo Analysis



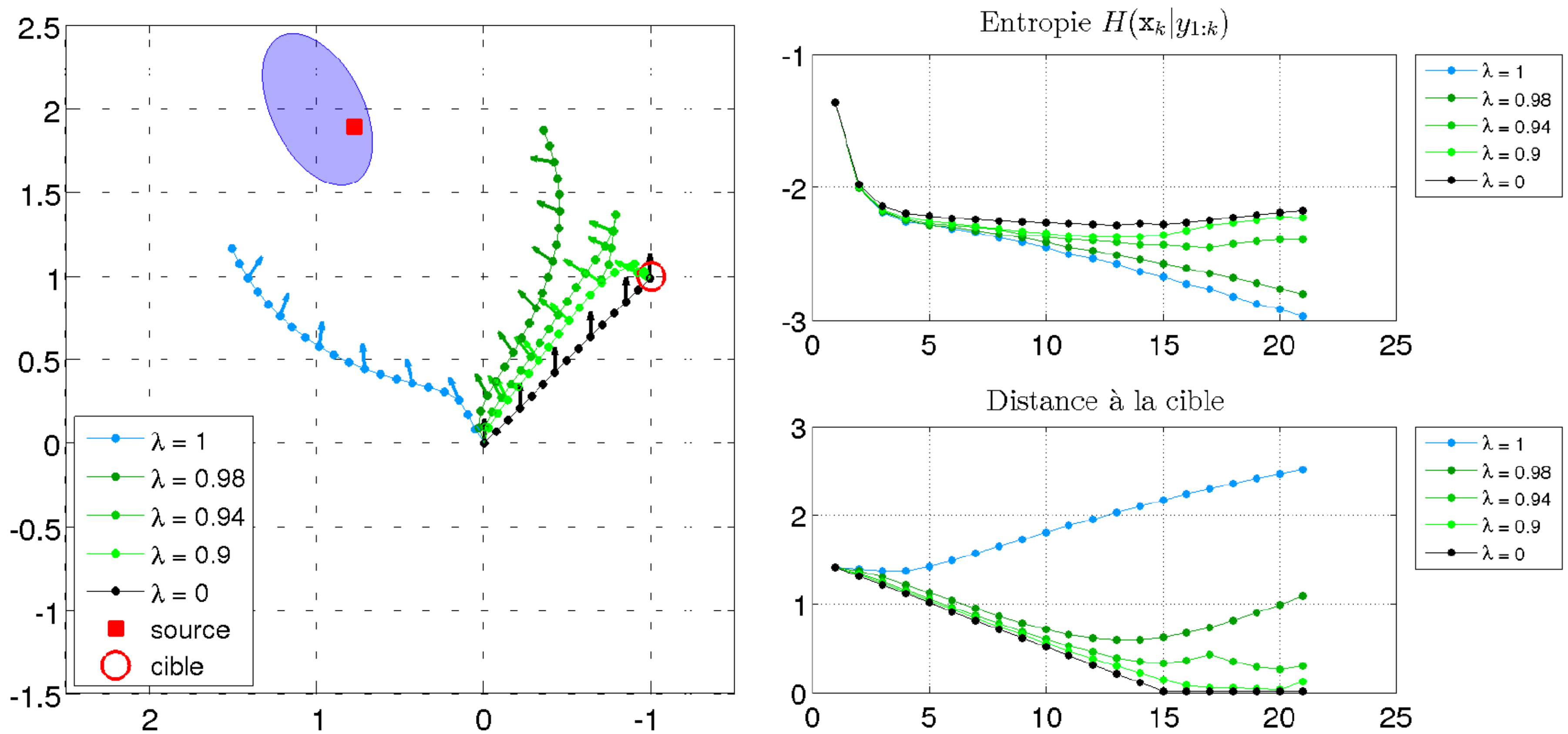


Conclusion

Optimal information-based feedback control for binaural localization

- Criterion: expected entropy of N-step-ahead state filtering pdf
- Unscented transform - Automatic differentiation - Projected gradient

- Other solutions do exist (e.g., MCTS of Nguyen, Vincent, Charpillet) but much farther to real-time
- Multi-objective criterion (e.g., exploration + servoing)



Prospects

PhD / PostDoc position can be opened by Fall 2019

- Incorporation of ML / stochastic / optimization methods to address
  - Multiple sources
  - Unanechoic/Dynamic environments
  - Incorporation of other sensing modalities
  - etc.

Thank You!

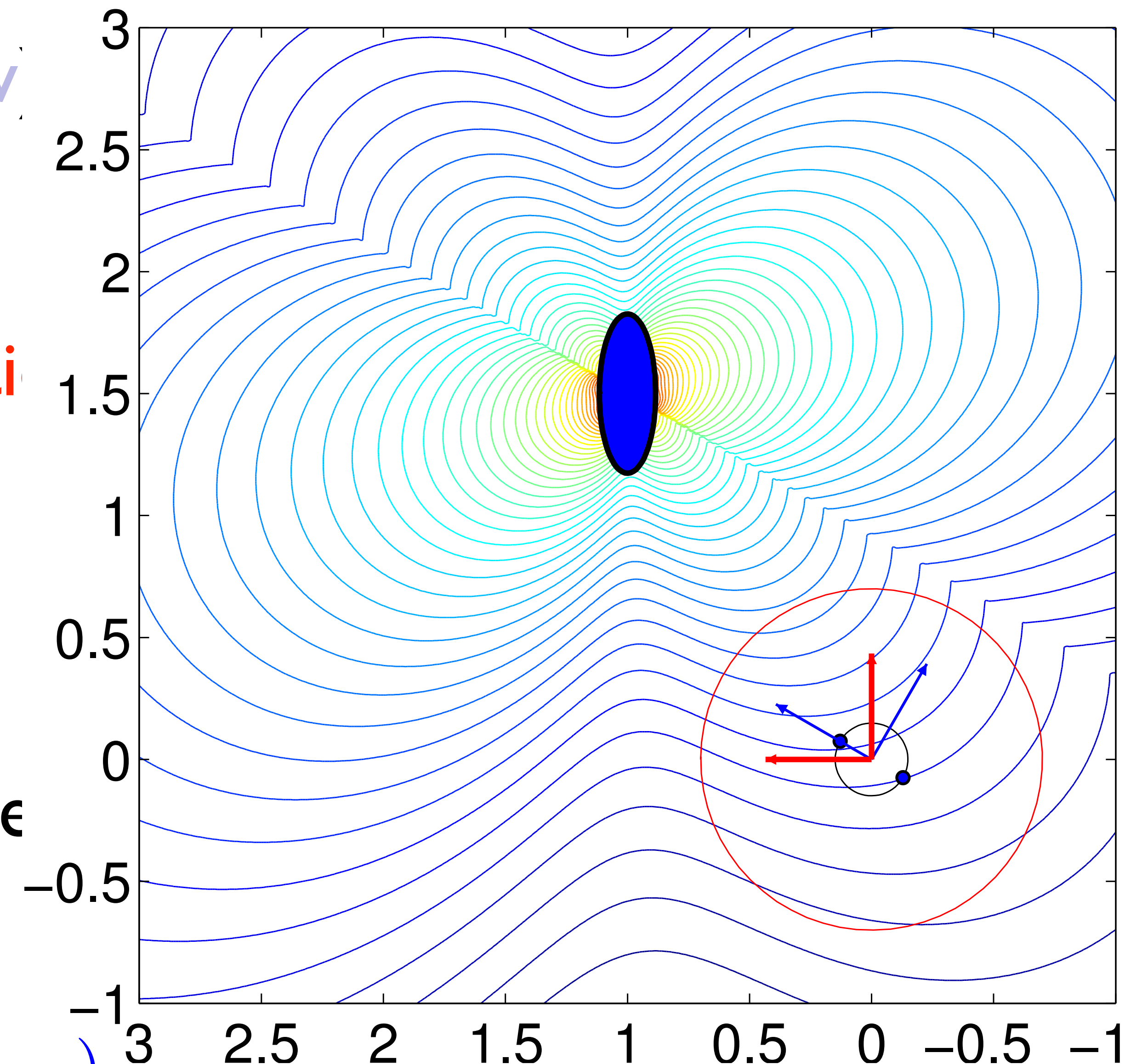
Patrick Danès

and many thanks to my colleagues!

One-step-ahead Solution
Contour Lines of the Criterion

Contour lines w.r.t translations for a given subsequent rotation

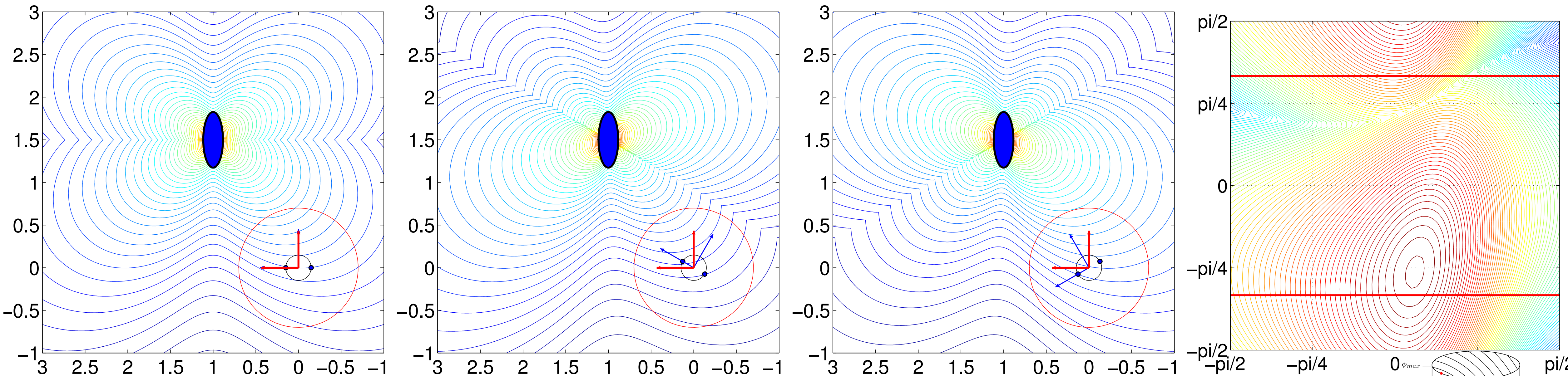
- Contour lines (warm=high,cold=low) of $F_k(T_y, T_z, \phi)$ w.r.t. T_y, T_z when followed by the rotation ϕ_0
- Red frame: sensor in the initial positi
- Red circle: $\mathcal{T}$
- Blue frame: orientation of the head if a zero translation were applied
- Initial estimate of the head-to-source position $\hat{x}_{k|k} = (1, 1.5)^T$ and 99%-probability confidence ellipsoid associated to $\mathcal{N}(x_k; \hat{x}_{k|k}, P_{k|k})$



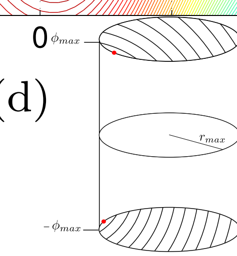
(b) $\phi_0 = +30°$

One-step-ahead Solution
Contour Lines of the Criterion

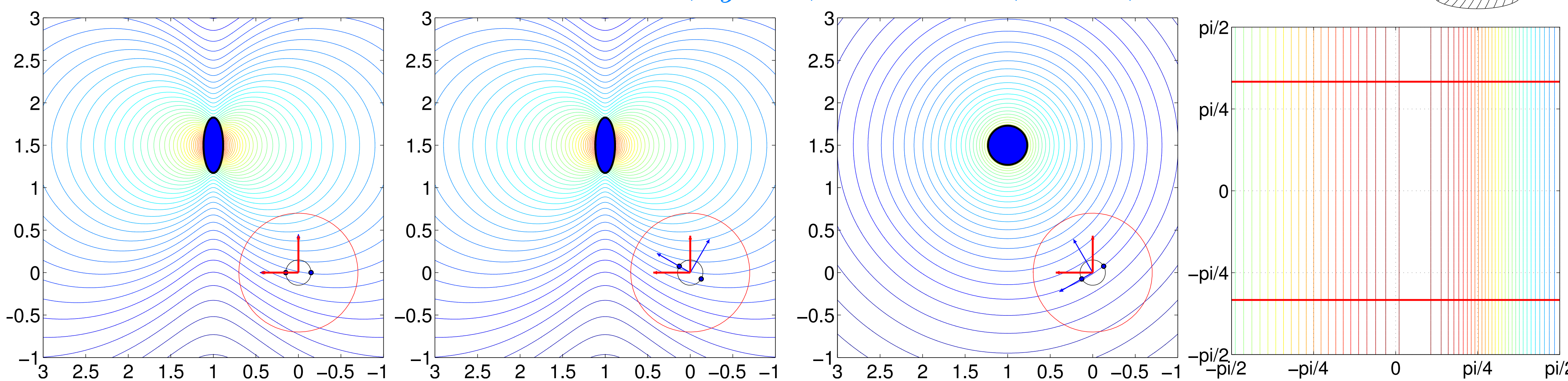
Remember that $z_k = \bar{l}(\theta_k) + v_k,\ \theta_k = -\mathrm{atan2}(e_y, e_z),\ v_k \sim \mathcal{N}(0, R_k)$



(a) $\phi_0 = 0$ (b) $\phi_0 = +30°$ (c) $\phi_0 = -30°$ (d)



What if $z_k = \theta_k + v_k,\ \theta_k = -\mathrm{atan2}(e_y, e_z),\ v_k \sim \mathcal{N}(0, R_k)$?



(e) $\phi_0 = 0$ (f) $\phi_0 = +30°$ (g) ϕ_0 has no impact (h)