Audio-based robot control from inter-channel level difference and absolute sound energy

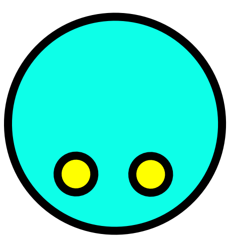
Aly Magassouba, Nancy Bertin and Francois Chaumette

IROS 2016

Presented by Youngki Kwon

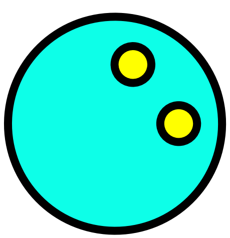
Motivation

- Goal : Positioning the robot with respect to the sound source at given distance and orientation.



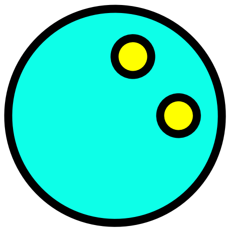
Motivation

- Goal : Positioning the robot with respect to the sound source at given distance and orientation.



Motivation

- Goal : Positioning the robot with respect to the sound source at given distance and orientation.



How?

Audio-based robot control

Audio-based robot control

- Real-time robot control to use sound!
- Aural Servo
  - Techniques which uses feedback information extracted from a aural sensor to control the motion of a robot

Geometric configuration

- Consider that the sound source is in front side of the robot
- X_s: omnidirectional sound source
- M_1, M_2: microphones
- M: midpoints of microphones
- R: center of robot
- $\dot{q}\ (u, w)$: control input (2-DOF)
  - u: velocity along $\overrightarrow{y_R}$
  - w: angular velocity around $\overrightarrow{z_R}$

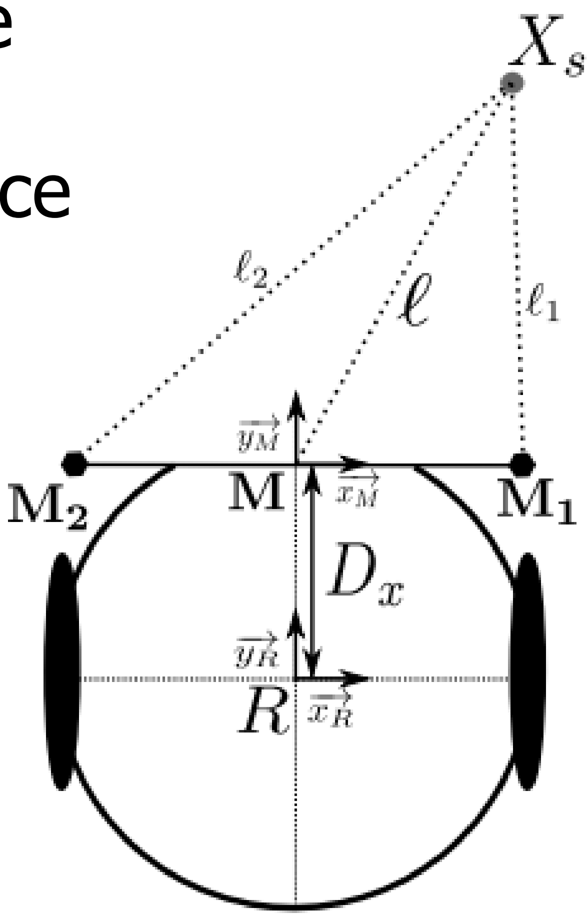
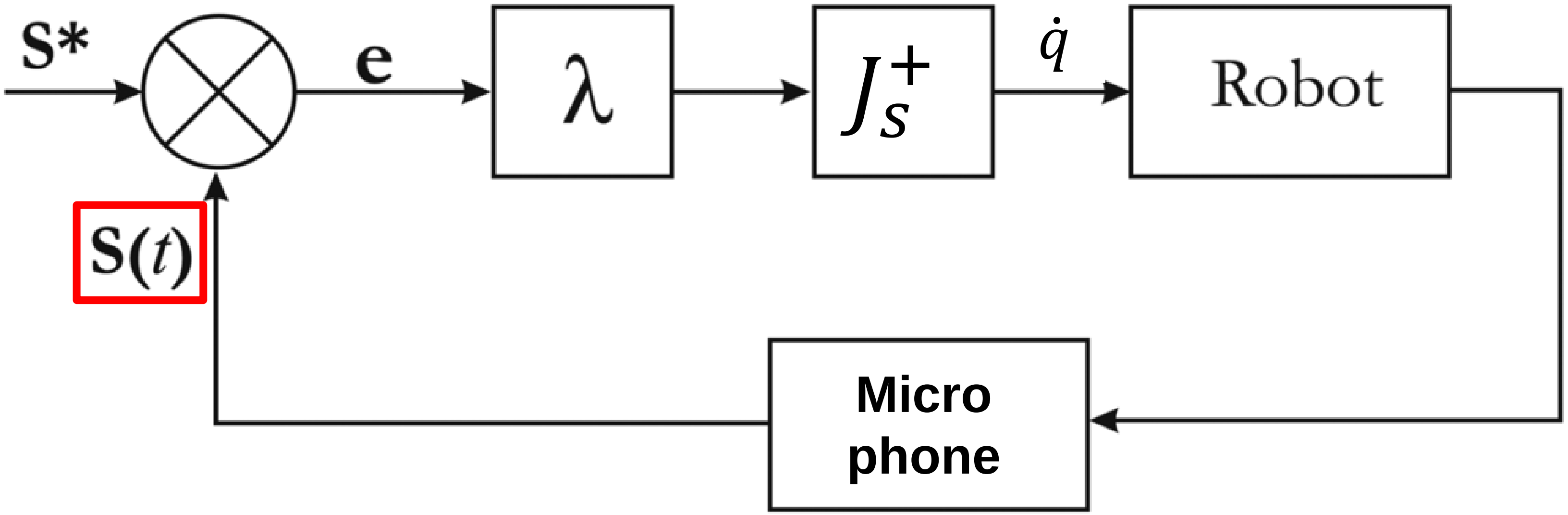


Fig. 1: Robot modelling

General framework

- The aim is to minimize an error $e(t)$

$$e(t) = s(t) - s^*$$



$$\dot{s} = L_s v, \ \dot{s} = J_s \dot{q} = L_s J_r \dot{q}, \ \dot{q} = -\lambda \widehat{J_s^+} e$$

$s(t)$: acoustic feature (e.g. ILD, Sound energy)

s^* : desired acoustic feature

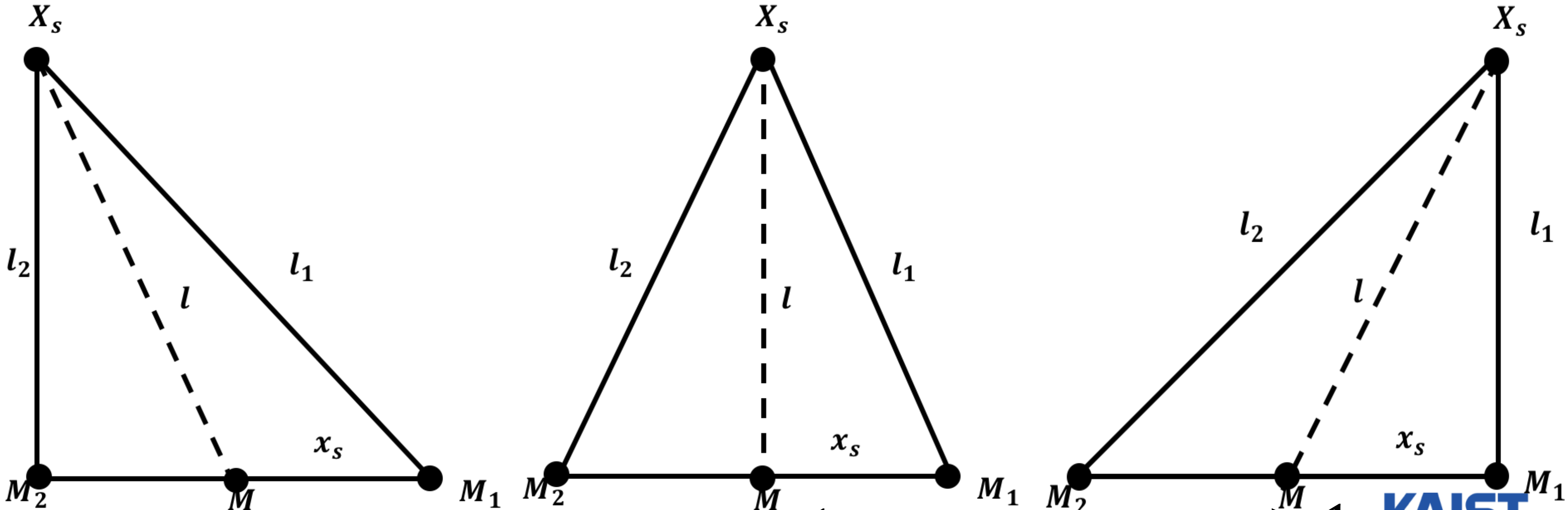
J_s^+ : Pseudo-inverse of interaction matrix

$\dot{q}$: control input

ILD based Aural Servo

- Inter-channel Level Difference (ILD)
  - Ratio of two amounts of energy received by each microphones
  - ILD ρ is approximated as $\frac{\ell_2^2}{\ell_1^2}$

$$\rho = \frac{E_1}{E_2} \approx \frac{\ell_2^2}{\ell_1^2}$$



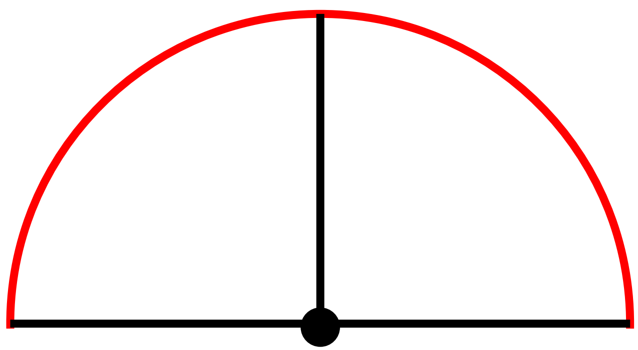
$\rho < 1$ $\rho = 1$ $\rho > 1$

ILD based Aural Servo

- Approximating position of sound source X_s
  - Condition :

$$E_1 \ell_1^2 = E_2 \ell_2^2.$$

$$(x_s - c_x)^2 + y_s^2 - \frac{E_1 E_2 d^2}{(E_1 - E_2)^2} = 0$$



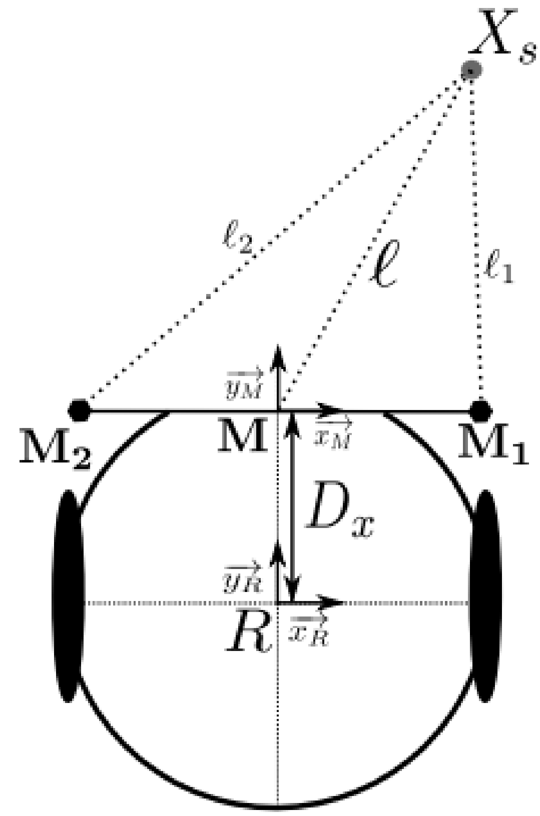
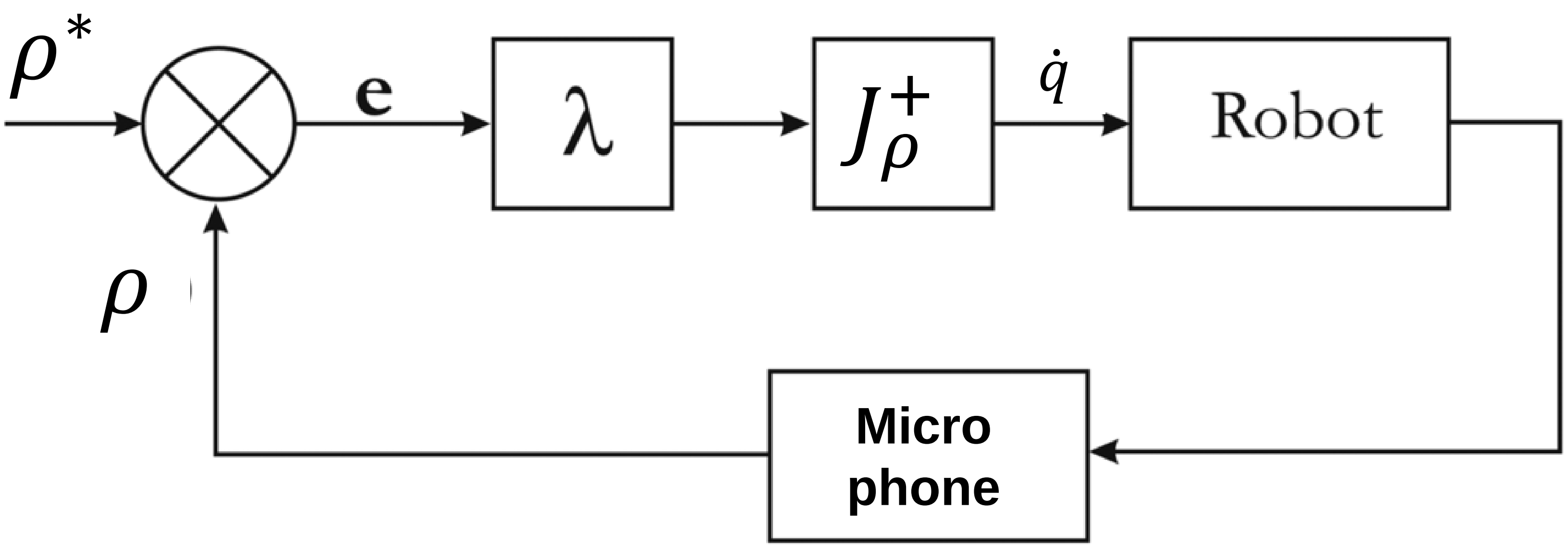


Fig. 1: Robot modelling

ILD based Aural Servo

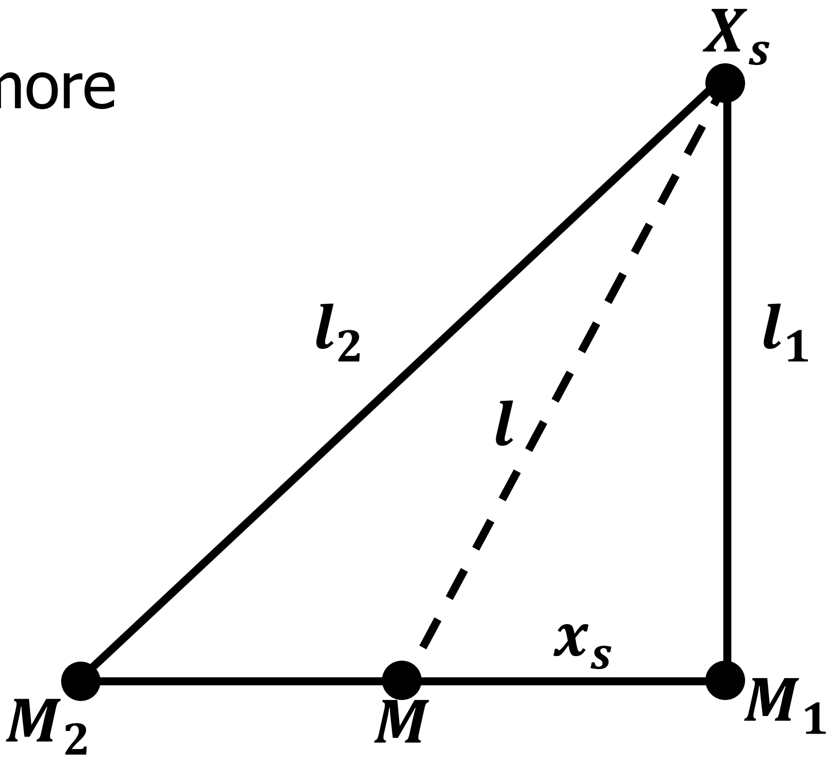
- Inter-channel Level Difference (ILD)
  - Ratio of two amounts of energy received by each microphones
  - ILD ρ is approximated as ${l_2^2}/{l_1^2}$

$$e(t) = \rho(t) - \rho^*$$



ILD accuracy limitation

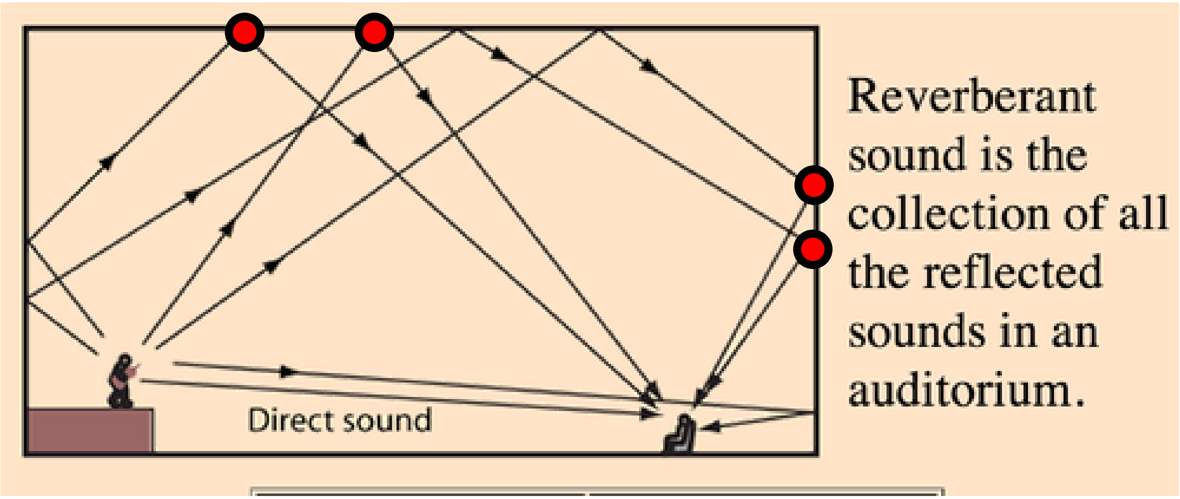
- When robot is far from sound source
  - $l_2^2 = l_1^2 + (2x_s)^{\ 2}$
  - $l\ \rightarrow\ \infty,\ l_1 \approx l_2,\ \rho \rightarrow 1$
  - ILD ρ is not significant anymore



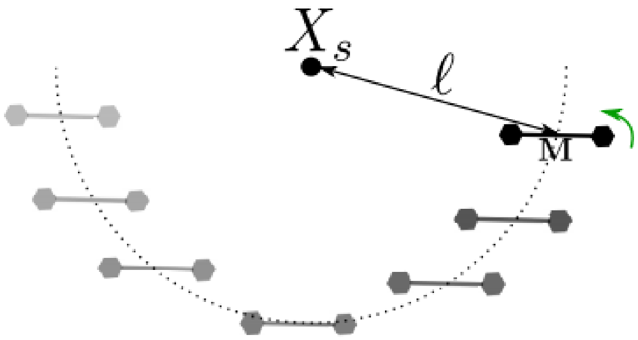
ILD accuracy limitation

- When reverberation time is high
  - Each microphones M_i perceive an additional energy from virtual sound sources
  - Make error e biased

$$\rho = \frac{\frac{1}{\ell_1^2} + \sum_{j=1}^{p} \frac{1}{\ell_{1_j}^2}}{\frac{1}{\ell_2^2} + \sum_{j=1}^{p} \frac{1}{\ell_{2_j}^2}}. \qquad e = \frac{\frac{1}{\ell_1^2} + e_{1p}}{\frac{1}{\ell_2^2} + e_{2p}} - \rho^*.$$



ILD accuracy limitation



RT_{60} (s)	ℓ (m)			
	0.5	**1**	**2**	**3**
0	< 1	< 1	2.7	6.4
0.05	< 1	3.6	5.6	7.6
0.1	< 1	5.1	7.0	8.4
0.2	< 1	8.8	10.4	21.4

Fig. 3: A simulated task that consists in orienting the robot in the direction of the sound source, from different poses in a 8×6 m^2 room. The final mean error, in degree, is calculated for several reverberation rate (RT_{60}) and distances to the source.

Accuracy is high only near by the sound source

+ Absolute sound energy

- Control the distance to the sound source by setting a desired energy level
- Reverberation has a minor effect when robot is nearby the sound source

$$E_{\mathrm{M}} = \frac{1}{\ell^2} \int_{t=0}^{w} a^2(t - \frac{\ell}{c})\,\mathrm{d}t.$$

$$e_{\mathrm{M}} = \left(\frac{1}{\ell^2} + e_p\right)\int_{t=0}^{w} a^2(t - \frac{\ell}{c})\mathrm{d}t - \left(\frac{1}{\ell{*2}} + e_p^*\right)\int_{t=0}^{w} a^2(t - \frac{\ell^*}{c})\mathrm{d}t \tag{28}$$

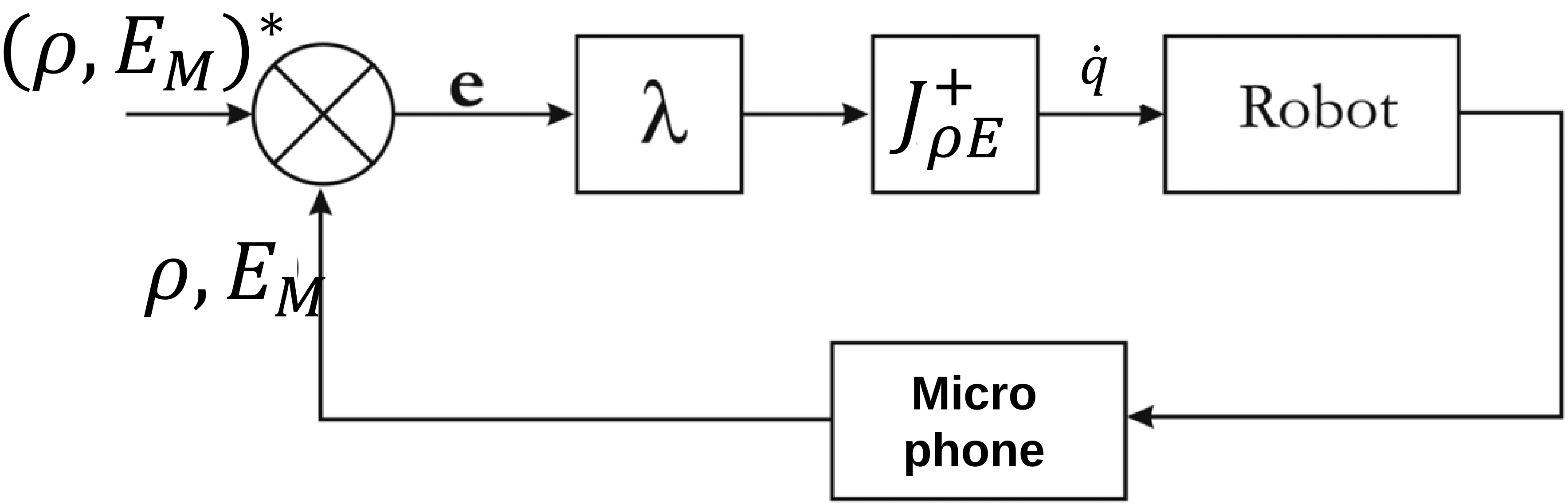
$$E_{\mathrm{M}} = \frac{E_1 + E_2}{2}.$$

KAIST

ILD + Absolute sound energy

- The ILD constrains the orientation of the microphones while the distance is constrained by the energy level

$$e(t) = (\rho, E_M)(t) - (\rho, E_M)^*$$



Experiment

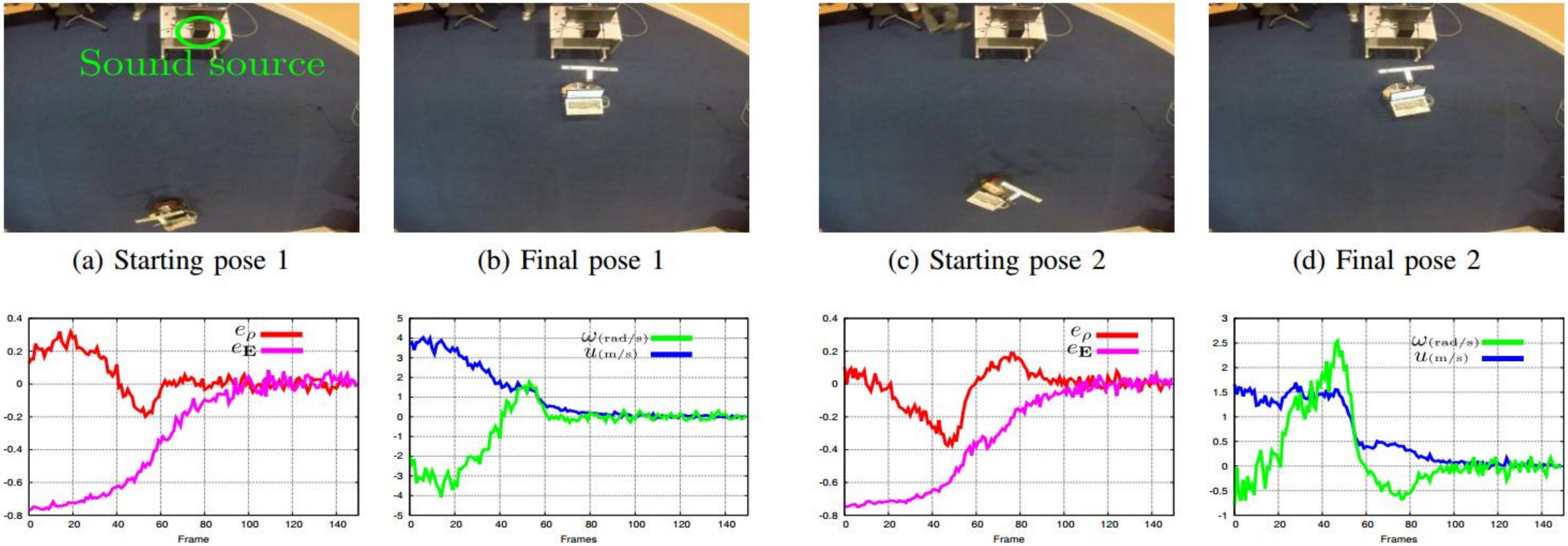
- Pioneer 3DX + Two omnidirectional microphones
- Three experiment scenarios
  - Typical positioning tasks
  - Long range navigation
  - Cooperative application

d	0.31 m
D_x	0.3 m
$\widehat{y}_s$	1 m
$\widehat{x}_s$	$\text{sign}(\rho - 1) \times 1$ m
$\lambda(x)$	$0.45e^{(-1.5x)}$

Fig. 6: Experimental settings

Typical positioning tasks

- $RT_{60} \approx 580ms, SNR = 20dB$, static sound source
- Desired acoustic feature : $l^* = 80cm, \rho^* = 1$



(a) Starting pose 1 (b) Final pose 1 (c) Starting pose 2 (d) Final pose 2

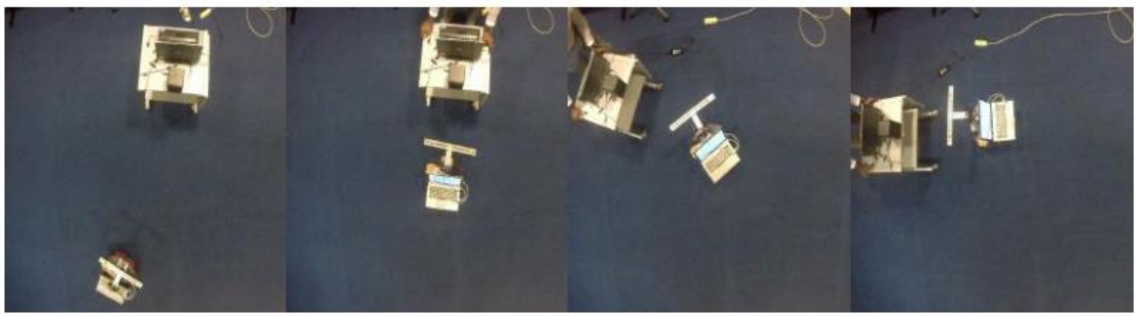
(e) Features error (f) Velocities input (g) Features Error (h) Velocities input

Fig. 5: Typical positioning tasks from two different starting poses

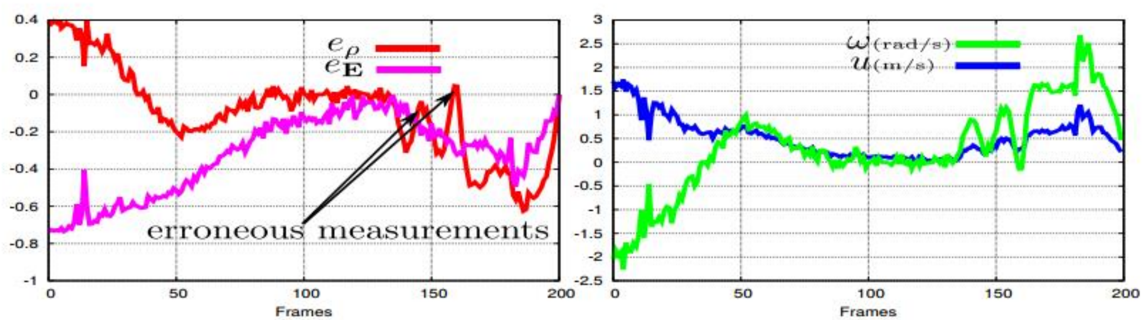
KAIST

Typical positioning tasks

- $RT_{60} \approx 580ms, SNR = 20dB$, dynamic sound source
- Desired acoustic feature : $l^* = 80cm, \rho^* = 1$



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



(e) Features error (f) Velocities input

Fig. 7: Following a moving sound source

Long range navigation

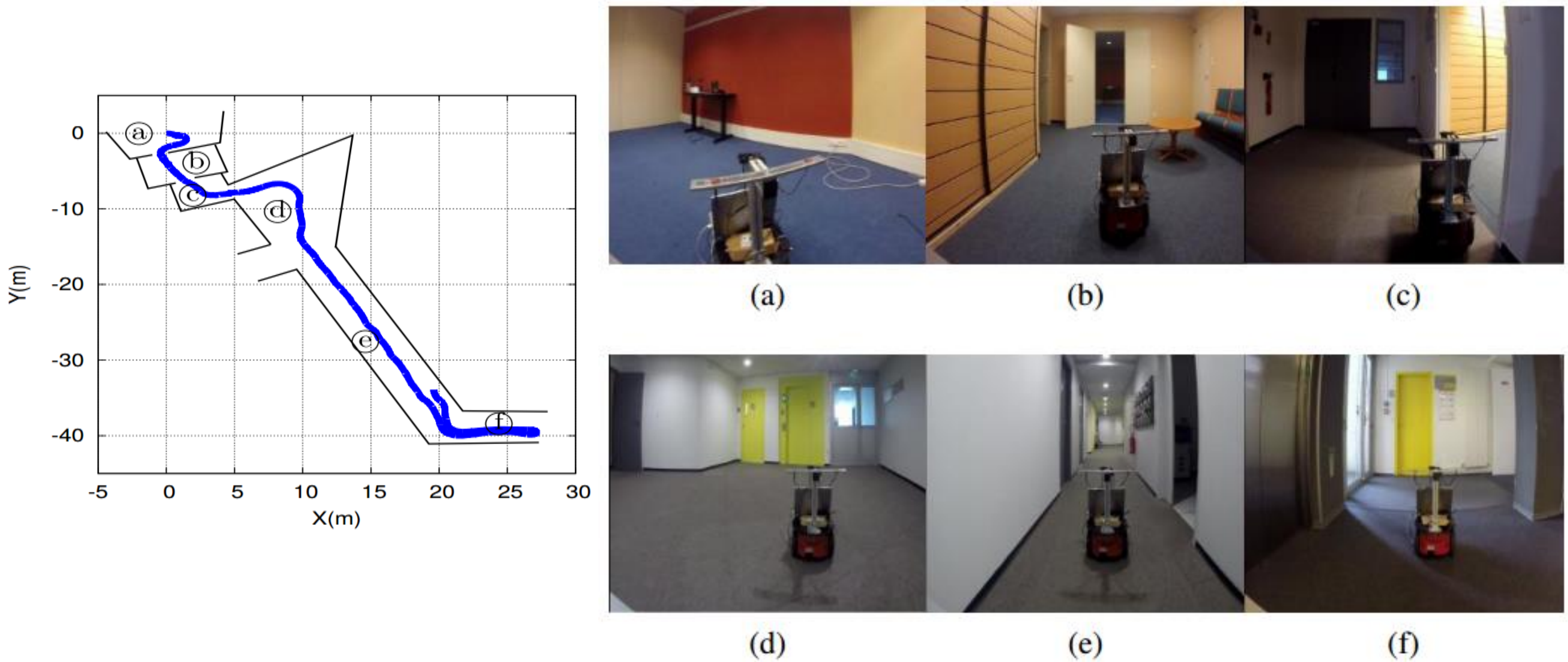
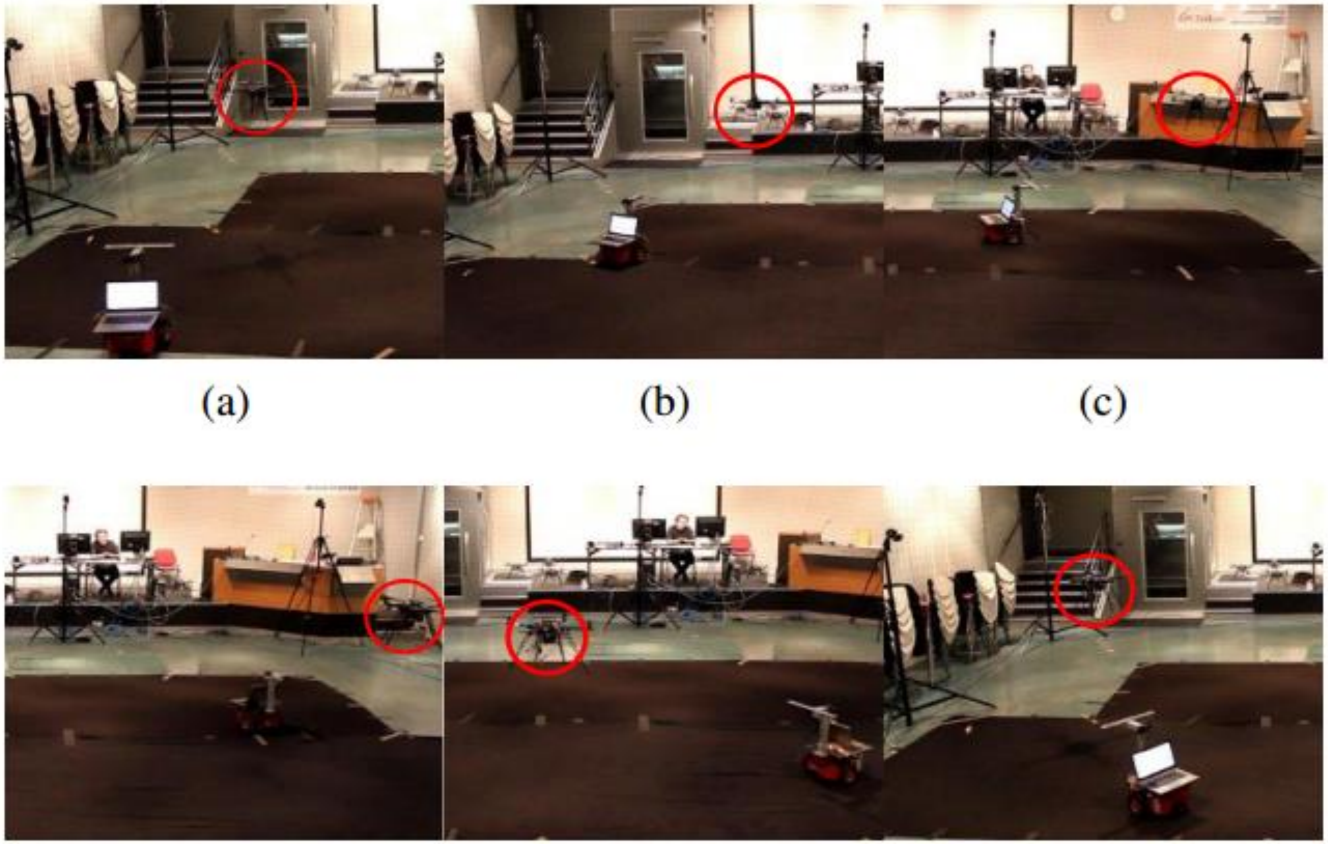


Fig. 8: Odometry data from the navigation task in indoor environment.The acoustic conditions for each location are respectively: ⓐ $RT_{60} \approx$ 580ms SNR$\approx$ 20dB, ⓑ $RT_{60} \approx$ 620ms SNR$\approx$ 20dB, ⓒ $RT_{60} \approx$ 680ms SNR$\approx$ 16dB, ⓓ $RT_{60} \approx$ 880ms SNR$\approx$ 13dB, ⓔ $RT_{60} \approx$ 700ms SNR$\approx$ 18dB, ⓕ $RT_{60} \approx$ 620ms SNR$\approx$ 17dB.

Cooperative application

- UAV led the unicycle ground robot by the sound from the propellers



(a) (b) (c) (d) (e) (f)

KAIST

Summary

- Audio-based robot control
  - Techniques which uses feedback information extracted from a aural sensor to control the motion of a robot
- Two acoustic features
  - Inter-channel Level Difference (ILD)
    - Finding direction of source
  - Sound Energy
    - Control distance to source