IROS2013 Robot Audition II

Outdoor Auditory Scene Analysis

Tokyo Big Site, November 5, 2013

Hiroshi G. Okuno

Professor, Ph.D, Fellow of IEEE and JSAI

Department of Intelligence Science and Technology

Graduate School of Informatics

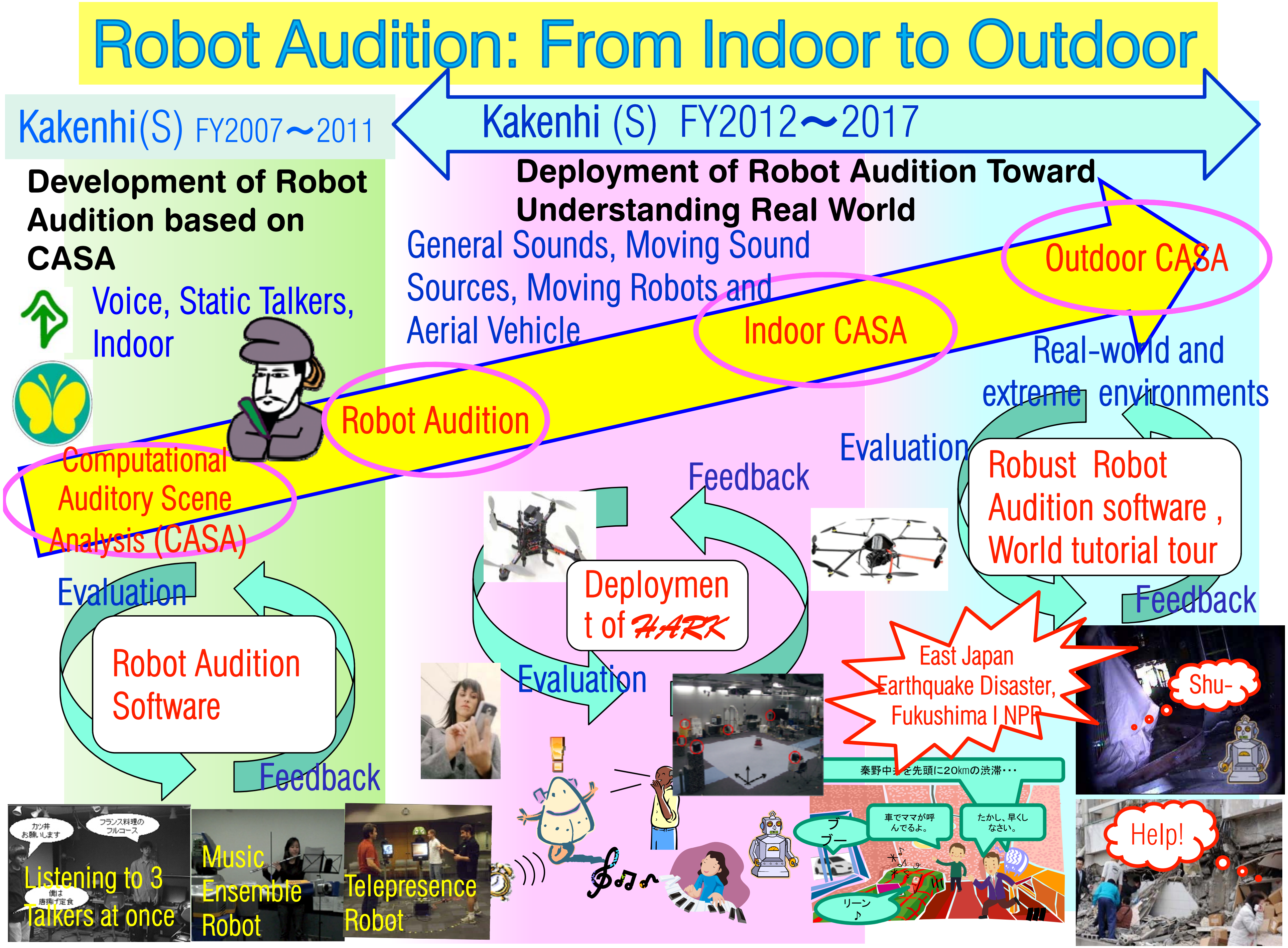
Kyoto University

http://winnie.kuis.kyoto-u.ac.jp/

okuno@i.kyoto-u.ac.jp

Outline of my talk

1. **Robot Audition so far**
   some demonstrations developed so far
2. **Motivations of Outdoor Auditory Scene Analysis**
   rescue robots and natural observations need listening capabilities, say, auditory scene analysis
3. **Current Status of Outdoor Auditory Scene Analysis**
   robot audition for animal acoustics, unmanned aerial vehicle (quadrocopter), and hose-typed rescue robots


Robot Audition: From Indoor to Outdoor
Kakenhi(S) FY2007～2011
Kakenhi (S) FY2012～2017
Development of Robot Audition based on CASA
Voice, Static Talkers, Indoor
Deployment of Robot Audition Toward Understanding Real World
General Sounds, Moving Sound Sources, Moving Robots and Aerial Vehicle
Computational Auditory Scene Analysis (CASA)
Robot Audition
Indoor CASA
Outdoor CASA
Real-world and extreme environments
Evaluation
Feedback
Robot Audition Software
Evaluation
Feedback
Deployment of HARK
Evaluation
Feedback
Robust Robot Audition software, World tutorial tour
East Japan Earthquake Disaster, Fukushima I NPP
Shu-
Help!
Listening to 3 Talkers at once
Music Ensemble Robot
Telepresence Robot
車でママが呼んでるよ。
たかし、早くしなさい。

Related Keynotes in Robot Audition

1. **Wrapping Up BINAAHR Project (Binaural Active Audition for Humanoid Robots)**
   Patrick Danes (LAAS-CNRS), TuBT8-1
2. **Introduction to HARK 2.0 – Open Source Software for Robot Audition**
   Kazuhiro Nakadai (Honda Research Institute Japan/ Tokyo Institute of Technology), TuDT8-1
3. **Map Generation and Scene Analysis for Robots**
   Satoshi Kagami (AIST), TuDT8-2

Some Outcome of Robot Audition

1. **Robot Audition Software *HARK***

   Open-sourced and free 10 tutorials, ego-noise cancellation, semi-blind separation, sound source localization and separation, non-parametric Bayesian signal processing

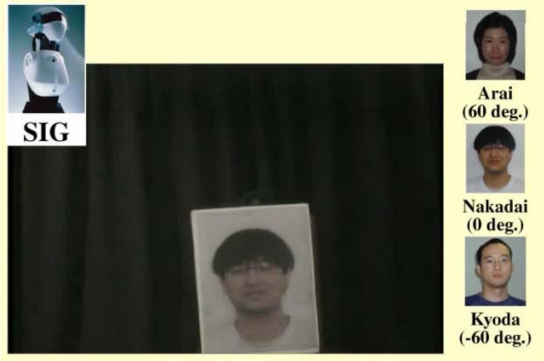
2. **Telepresence Robot**

   Visualization of *HARK* Output based on visual information seeking mantra [Schneidermann]

3. **Music Co-player Robots**

   Theremin players and ensembles with human players

Progress of Simultaneous Listening

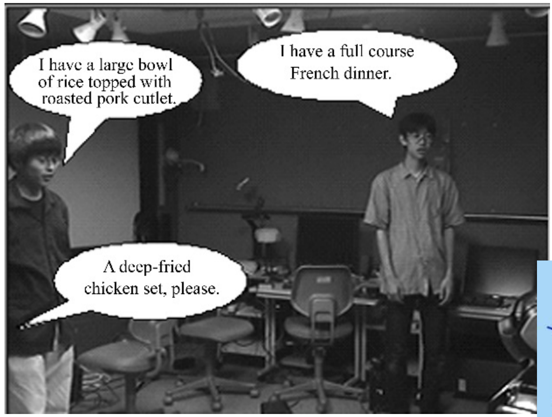


In 2002

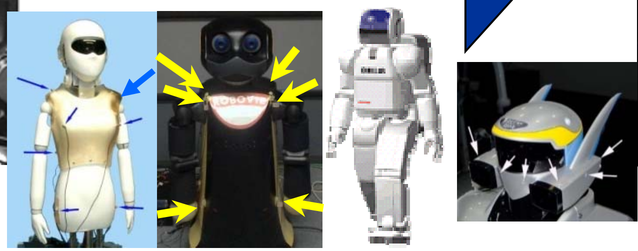
narrower interval between speakers

In 2003

- actual human speakers
- sentence recognition
- in a larger room
- less *prior* information
- on 3 kinds of robots



In 2005



4 times speed-up by FlowDesigner

In 2006

2003: 3 Speakers at 30° Int'val

2006: speed up by *HARK*

Response time 1.9 sec on *HARK*

Some Outcome of Robot Audition

1. **Robot Audition Software *HARK***
   Open-sourced and free 10 tutorials, ego-noise cancellation, semi-blind separation, non-parametric Bayesian signal processing, Sound source localization and separation

2. **Telepresence Robot**
   Visualization of *HARK* Output based on visual information seeking mantra [Schneidermann]

3. **Music Co-player Robots**
   Theremin players and ensembles with human players

HARK on Telepresence Robot

1. Telepresence Robot **Texai** under development at Willow Garage (Menlo Park, CA, USA)

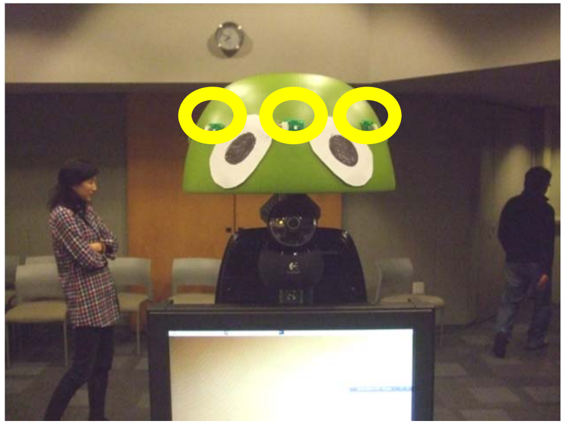
   http://www.willowgarage.com/

2. 2D visualizer for HARK output based on "**visual-information seeking mantra**" [Schneidermann] *"Overview first, zoom and filter, then details on demand"*

3. Install HARK on their Texai and develop a new interface for sound source localization and separation with sound focus control mechanism.

HARK on Telepresence robot

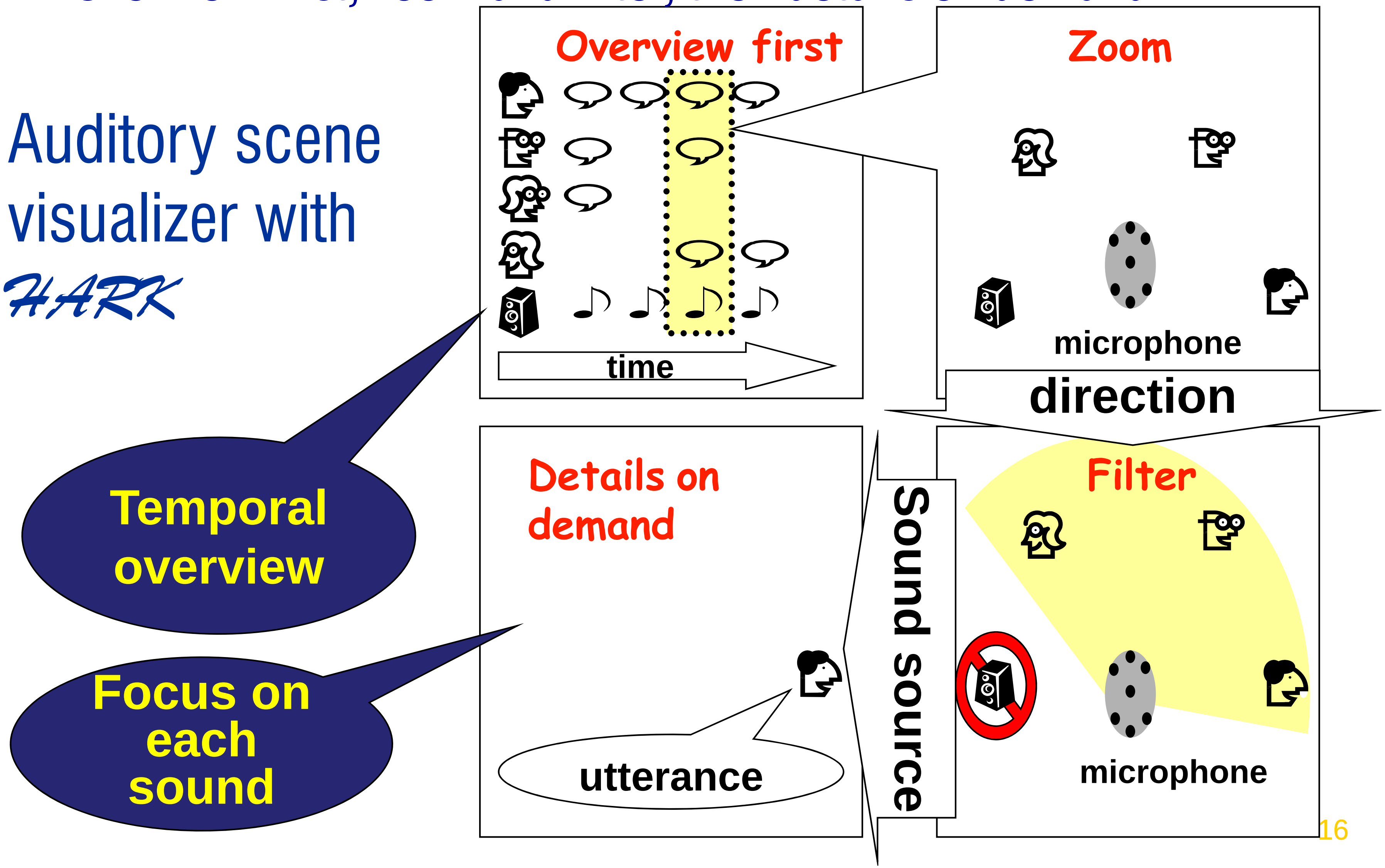
- **Texai**, Willow Garage's telepresence robots
- Head with 8 microphones is added.



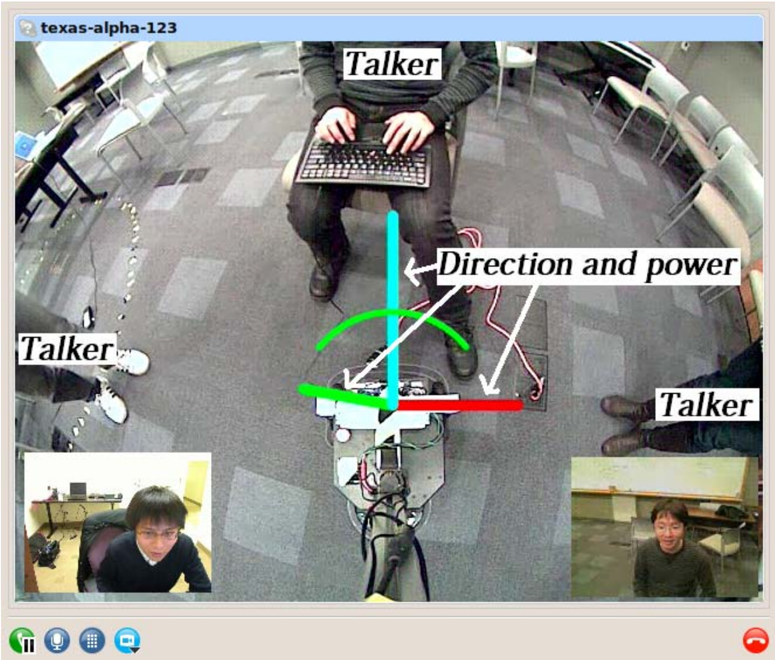
CASA Visualizer with HARK

Visual-information seeking mantra [Schneidermann]
"Overview first, zoom and filter, then details on demand"

Auditory scene visualizer with HARK



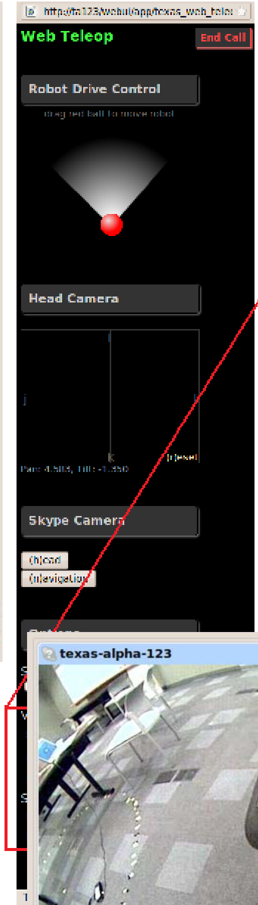
New functions based on HARK



Sound source localization results on the viewer

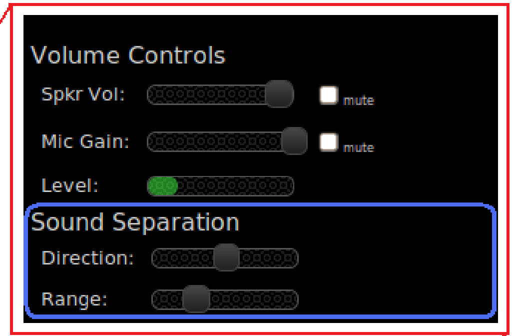
Line direction: Sound direction

Line length : Power

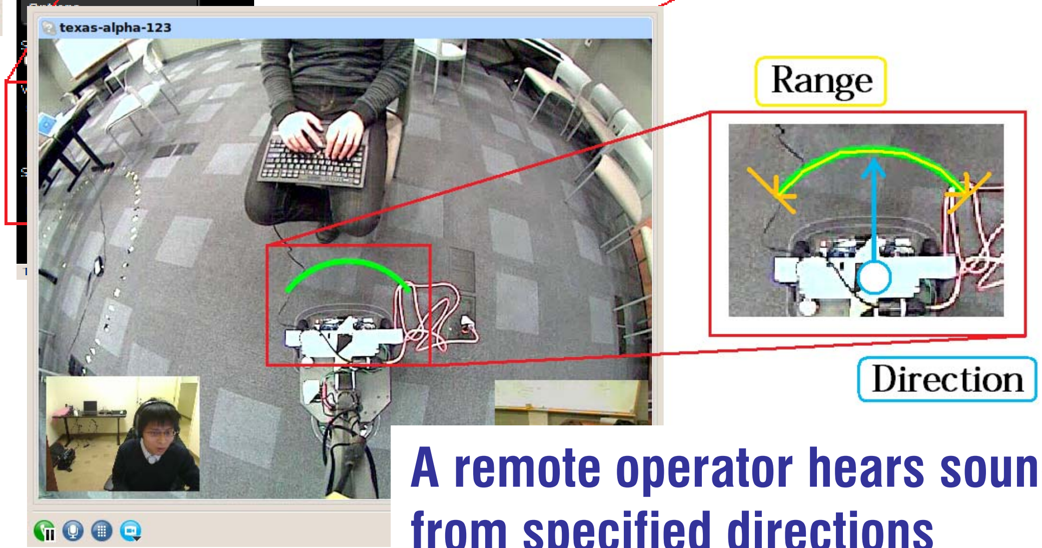


Direction-based sound filtering

Specify the center direction and range of the filter.



New operation



A remote operator hears sound from specified directions

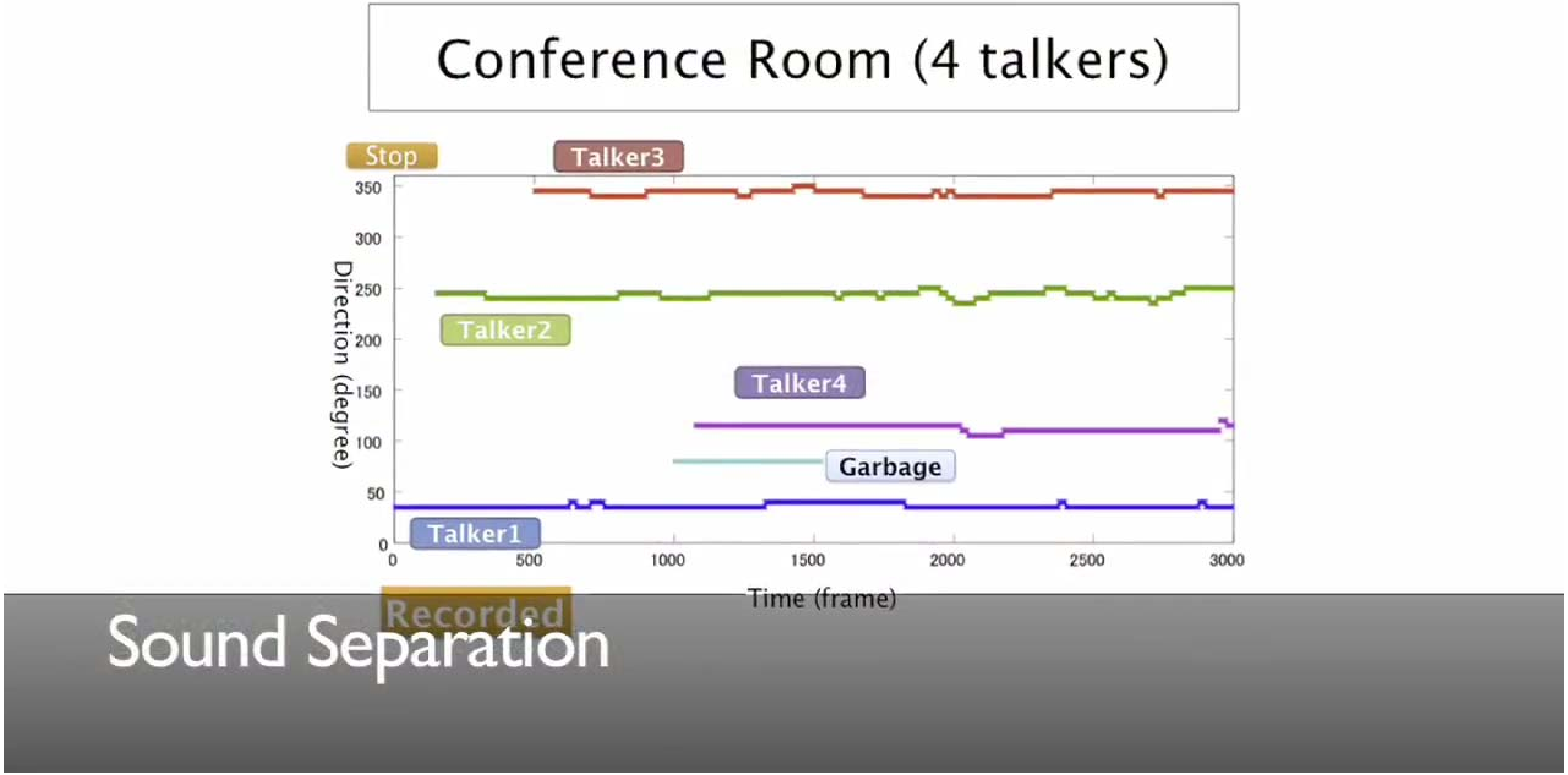
Demo: Texai with 3 people & Texai

From Indiana

Menlo Park, California, USA

http://www.willowgarage.com/

Demo: Texai with 3 people & Texai



http://www.willowgarage.com/

Some Outcome of Robot Audition

1. **Robot Audition Software *HARK***
   Open-sourced and free 10 tutorials, ego-noise cancellation, semi-blind separation, non-parametric Bayesian signal processing, Sound source localization and separation

2. **Telepresence Robot**
   Visualization of *HARK* Output based on visual information seeking mantra [Schneidermann]

3. **Music Co-player Robots**
   Theremin players and emsembles with human players

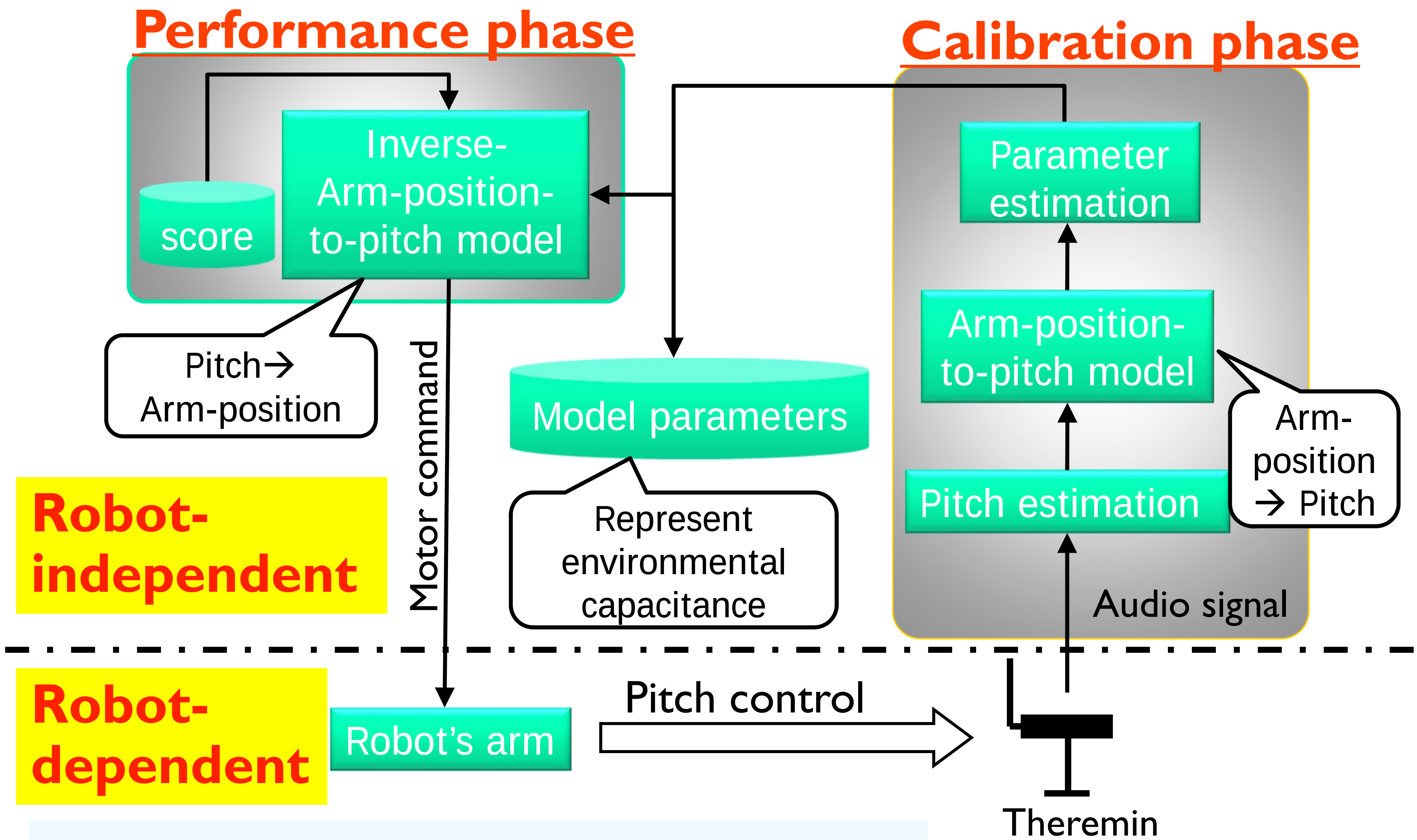
Human-Robot Interaction through Music

1. **Why through music?**
   An entertainment robot beyond the cultural barriers, e.g., generation, gender, country, race, ... unlike languages
2. **Active commitment in interaction**
   People can participate the entertainment

From audience (passive)

To a participant (active)

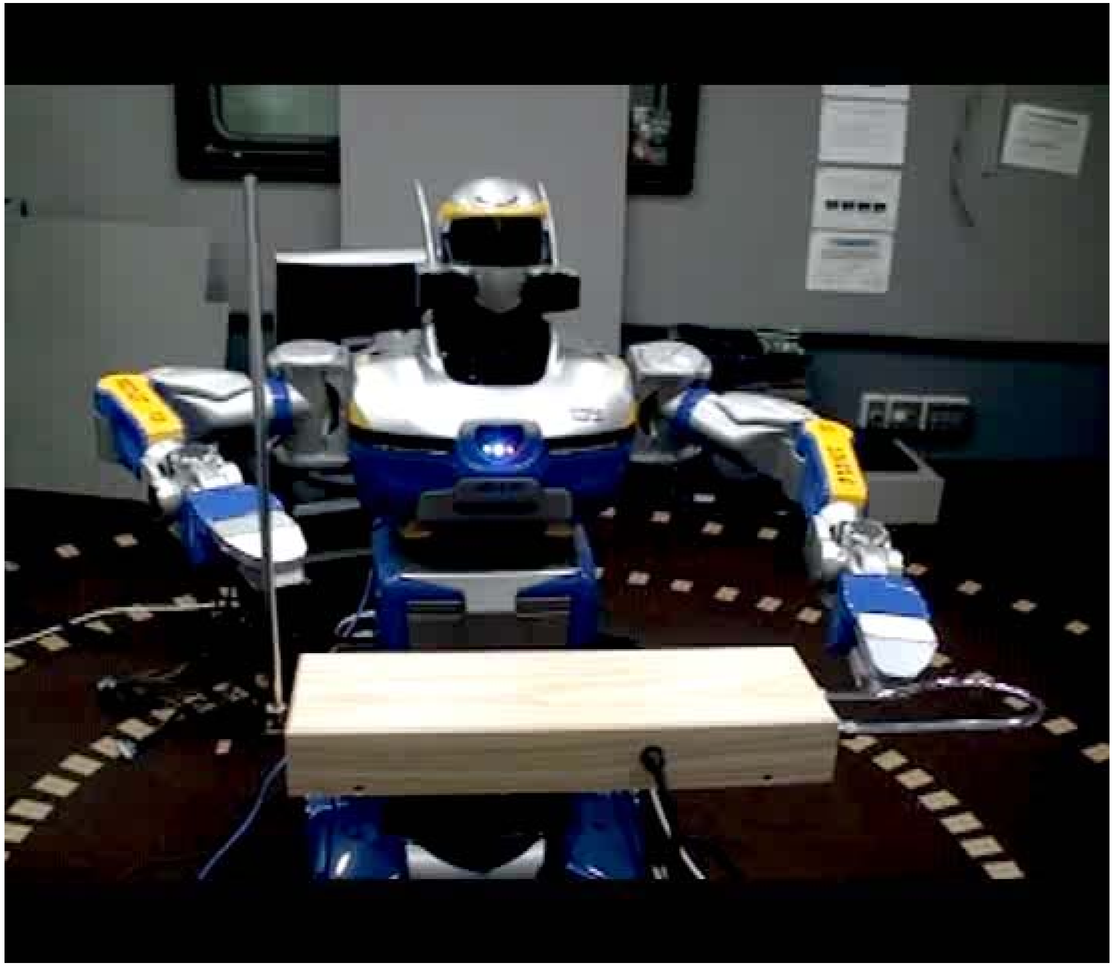
Thereminist Robot System



HRP-2, HERO, NAO, ASIMO, Hearbot

HRP-2 Plays the Theremin

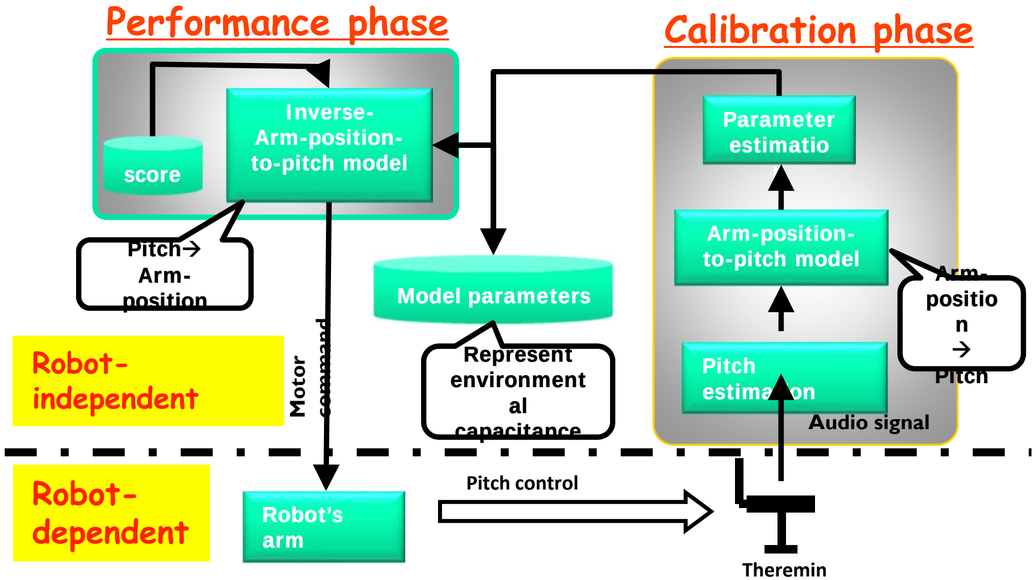
Volume control antenna

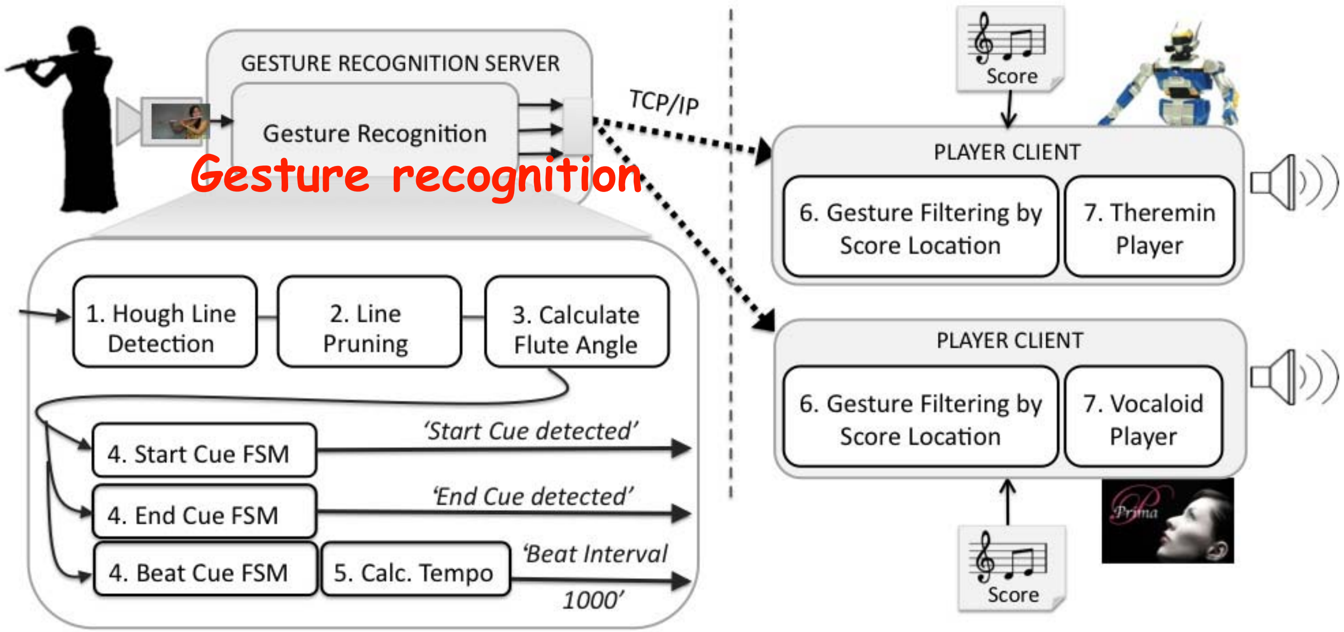


Pitch control antenna

Musical Robots for Ensemble

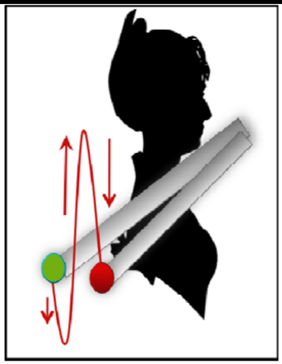
- Theremin Playing Robots
- *HARK*-Music
- Co-player robot for Ensemble



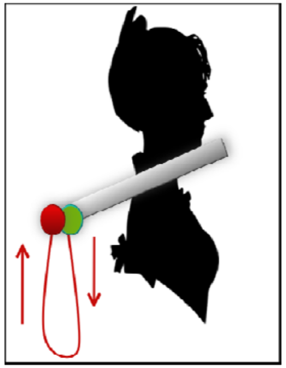


Musical Robots for Ensemble

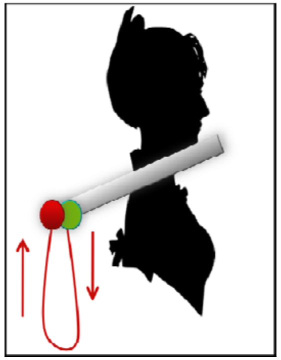
Start Cue



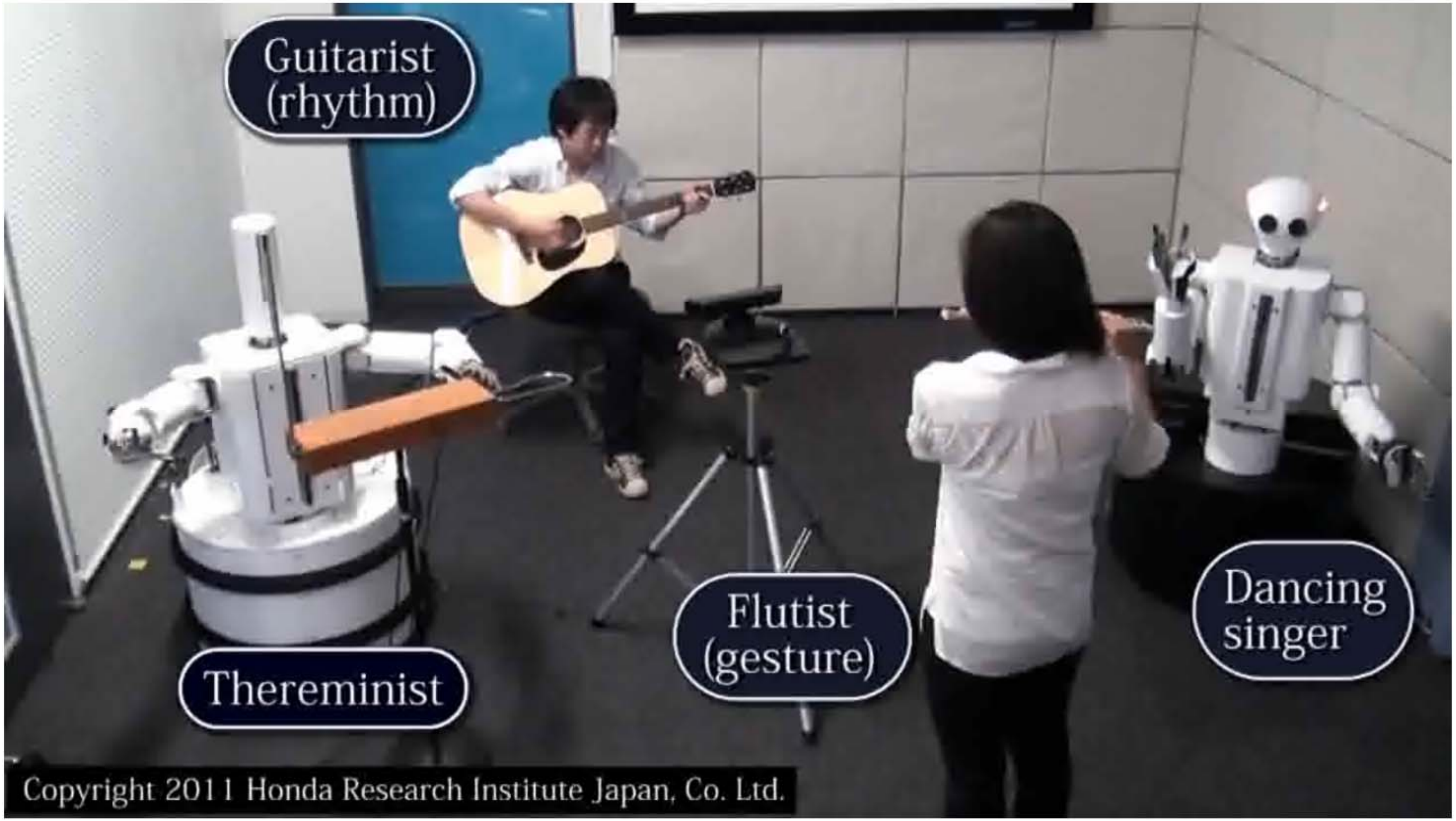
Beat Cue



End Cue



Quartet: 2 Robots and 2 humans



Audio-Visual Integration for Beat Tracking [IEEE Humanoids 2012]

TRIO: NAO Plays Theremin

Outline of my talk

1. **Robot Audition so far**
   some demonstrations developed so far
2. **Motivations of Outdoor Auditory Scene Analysis**
   rescue robots and natural observations need listening capabilities, say, auditory scene analysis
3. **Current Status of Outdoor Auditory Scene Analysis**
   robot audition for animal acoustics, unmanned aerial vehicle (quadrocopter), and hose-shaped rescue robots

Motivation: Outdoor Auditory Scene Analysis

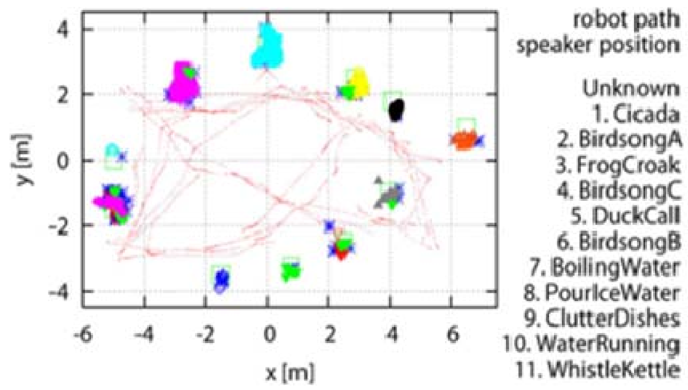
1. **CASA in indoor and outdoor/sky environments should be more robust**
   Feasible only in laboratory environments
2. **Rescue robots need listening capabilities**
   Rescue robots do not exploit the possibilities of listening.
3. **Advanced signal processing is needed by natural observations, communication of frogs, that of birds.**
   Robot audition is actually used for human-robot interactions, but only a few for other applications.

Issues: Outdoor Auditory Scene Analysis

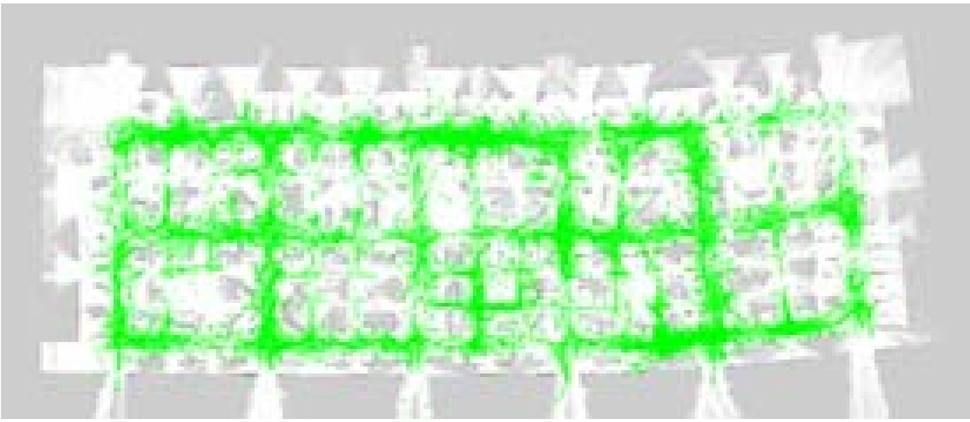
1. **Development of CASA technologies**
   Non-parametric Bayesian signal processing
   Sound activity detection (5W1H), Auditory map generation, visualization for sound awareness
2. **Listening from UAV**
   Sound source localization robust against motor noise and wind roar, dynamic calibration of UAV's unstable attitude, 3D auditory map, cooperation between land and sky
3. **Listening in natural and disastrous environments**
   Bird/frog recognition, bird/frog song recognition (love, territory, alarm, ..), analysis of bird song communication between different species (collision detection and avoidance like Ethernet), provide auditory awareness

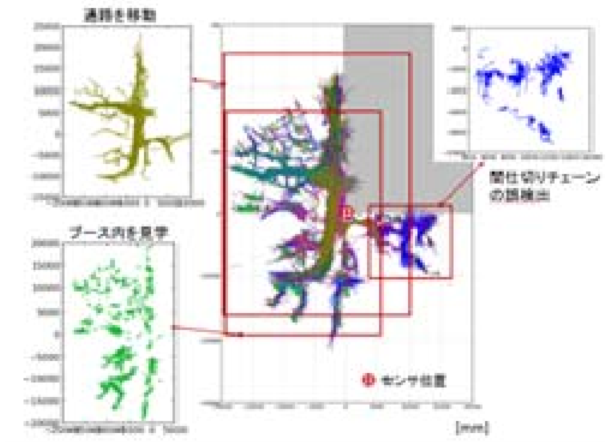
Auditory Map Generation

1. From CASA output to Auditory Map Generation



2. Simutaneous People tracking and Mapping

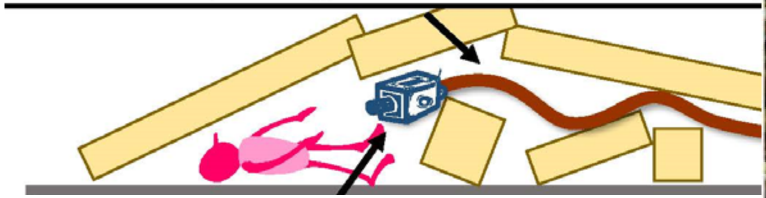




For Rescue work

Outdoor auditory scene analysis is essential.

- ***Useful for finding victims in a disaster situation***




Robot audition and CASA should provide*:*

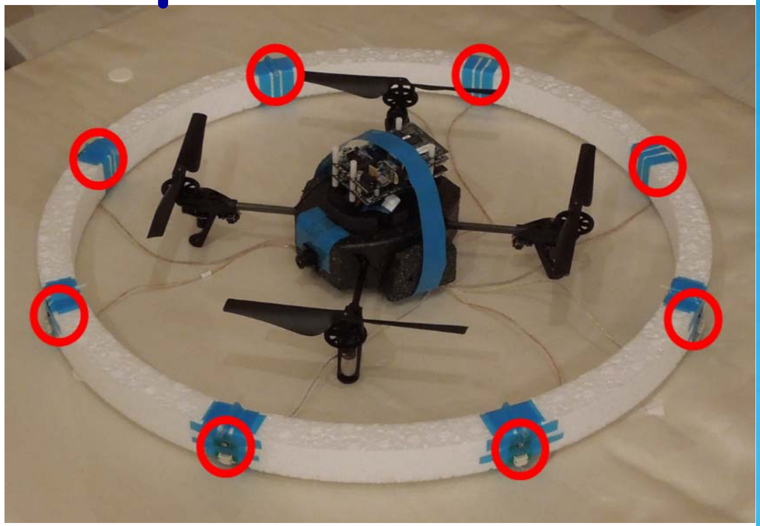
- ***Sound source localization, sound source separation, speech recognition, separated sound identification***

However, studied only in laboratory, indoor or simulated environments.

- **UAVs to capture sounds from sky**
- **Hose-shaped robots to capture sounds under debris**

Robot audition for rescue robots

Unmanned Aerial Vehicle
Hose-shaped Robots

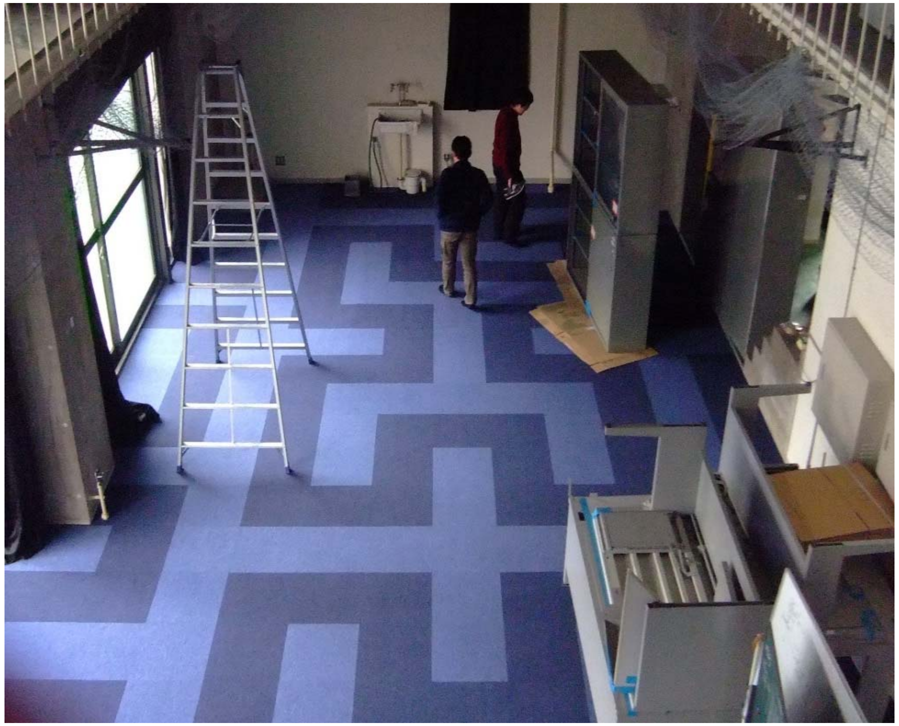


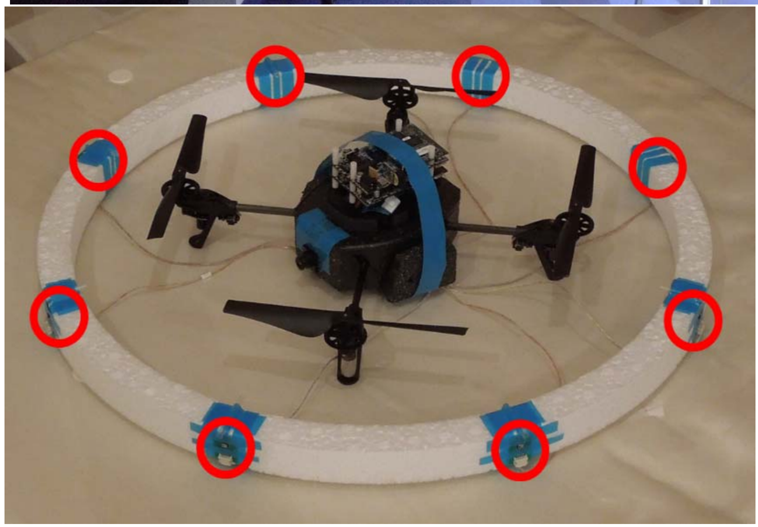
Active Scope Camera

Outline of my talk

1. **Robot Audition so far**
   some demonstrations developed so far
2. **Motivations of Outdoor Auditory Scene Analysis**
   rescue robots and natural observations need listening capabilities, say, auditory scene analysis
3. **Current Status of Outdoor Auditory Scene Analysis**
   robot audition for animal acoustics, unmanned aerial vehicle (quadrocopter), and hose-typed rescue robots

Experimental room for UAV





Indoor field for UAV

Quadrotor helicopter with auditory device

Autonomous UAV with microphone array

- Sensor readings, motor command and audio signal can be measured synchronously.
- Accurate localization (reference data) : RTK-GPS
- 1.4kg+ payload, 10min flight time

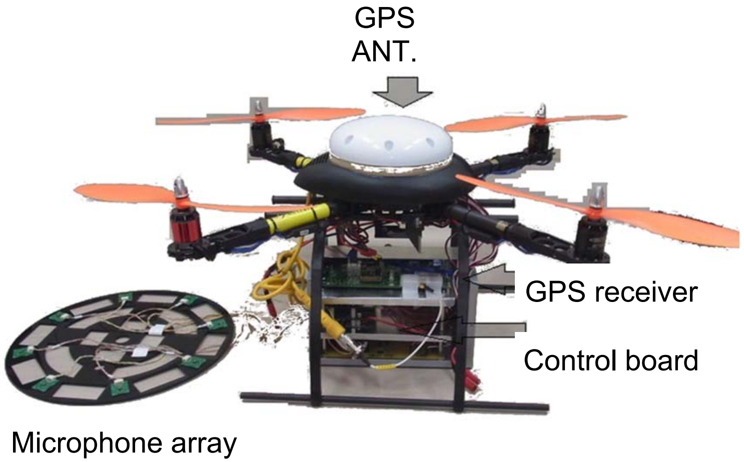
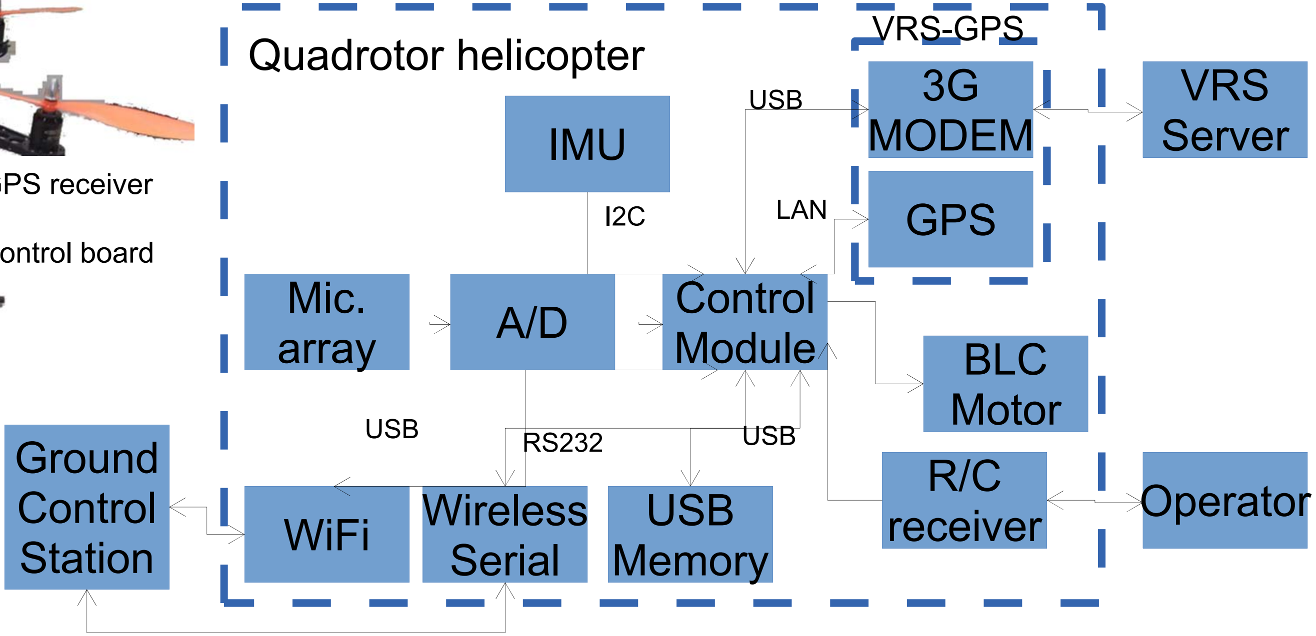


Photo of the developed quadrotor helicopter (above), and its system structure (right fig).



Quadrotor helicopter with auditory device

Autonomous UAV with microphone array

- Sensor readings, motor command and audio signal can be measured synchronously.
- Accurate localization (reference data) : RTK-GPS
- 1.4kg+ payload, 10min flight time

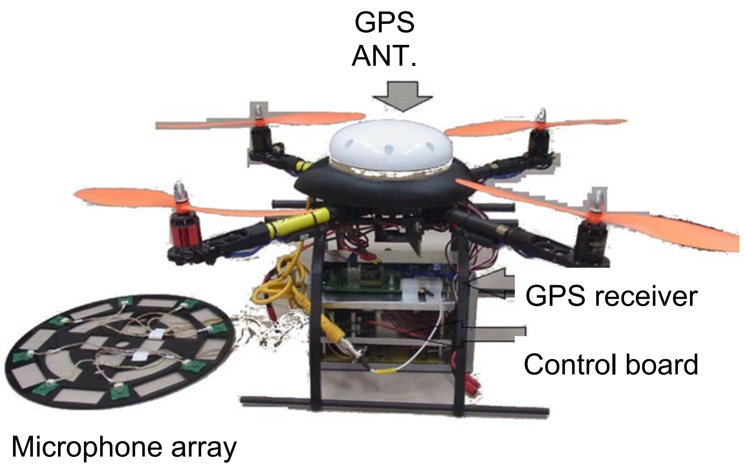
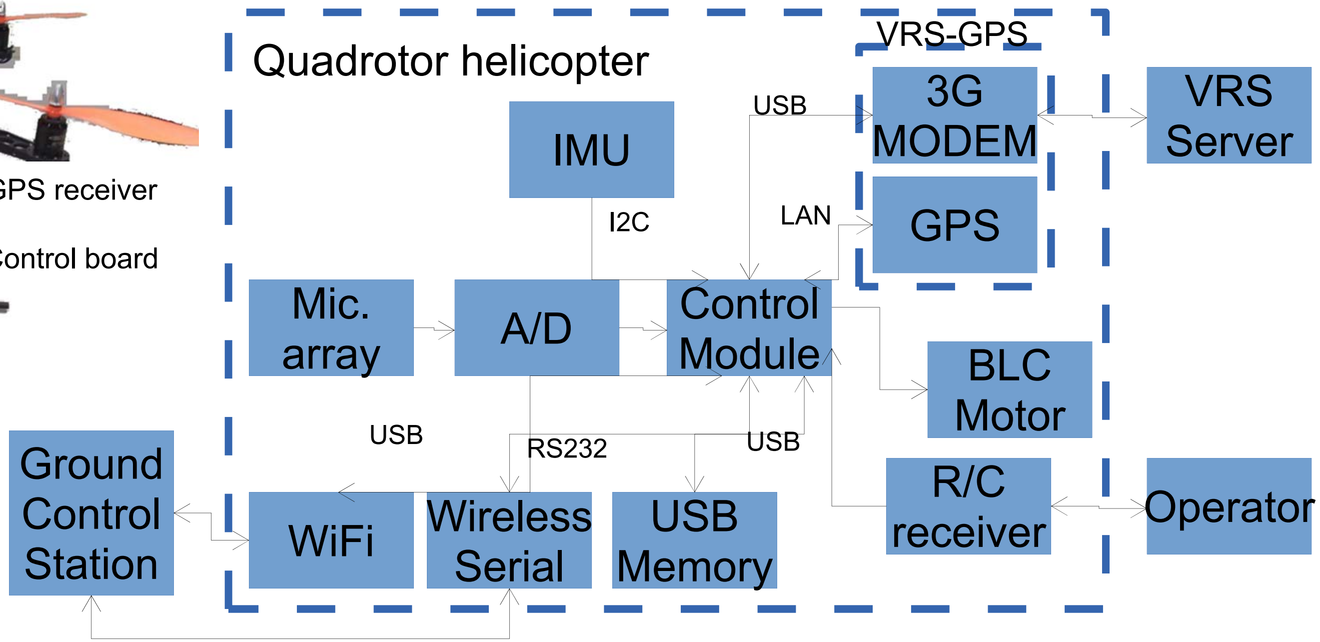
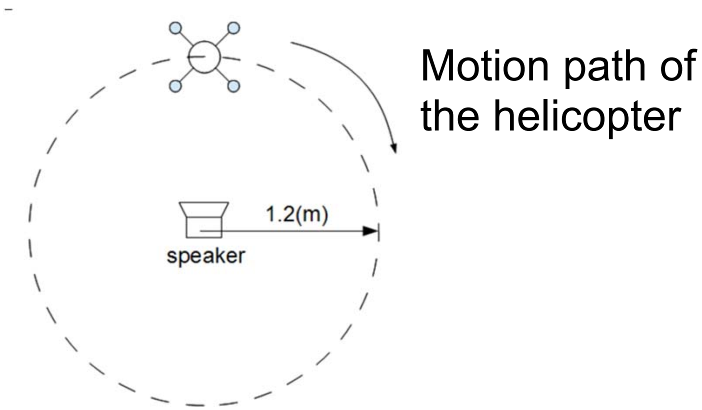


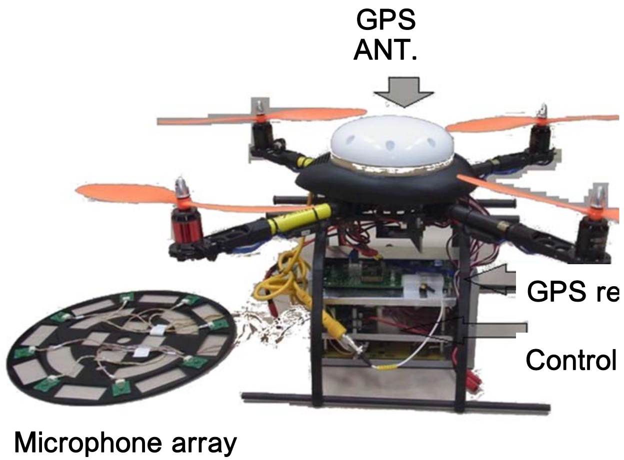
Photo of the developed quadrotor helicopter (above), and its system structure (right fig).

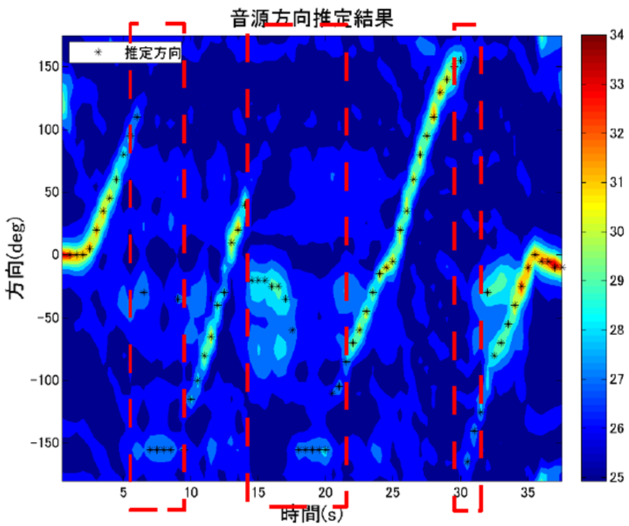


Quadrotor helicopter with auditory device

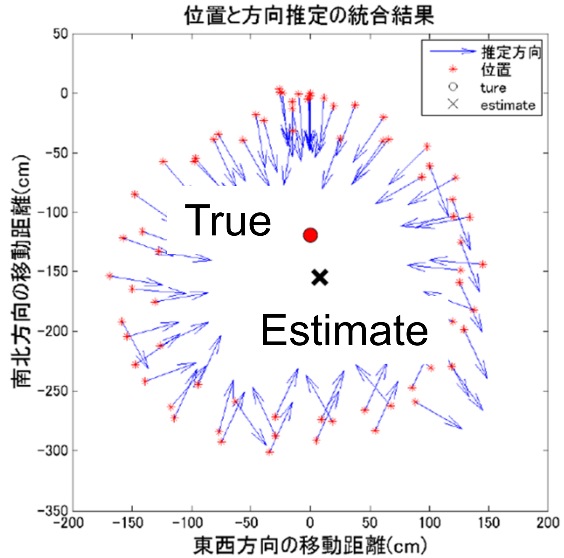
- Carry the helicopter around a speaker with the rotors rotating.
- Bearing to the speaker was estimated by MUSIC (*HARK*).
- Estimate the speaker location based on LMS estimation.







MUSIC spectrum

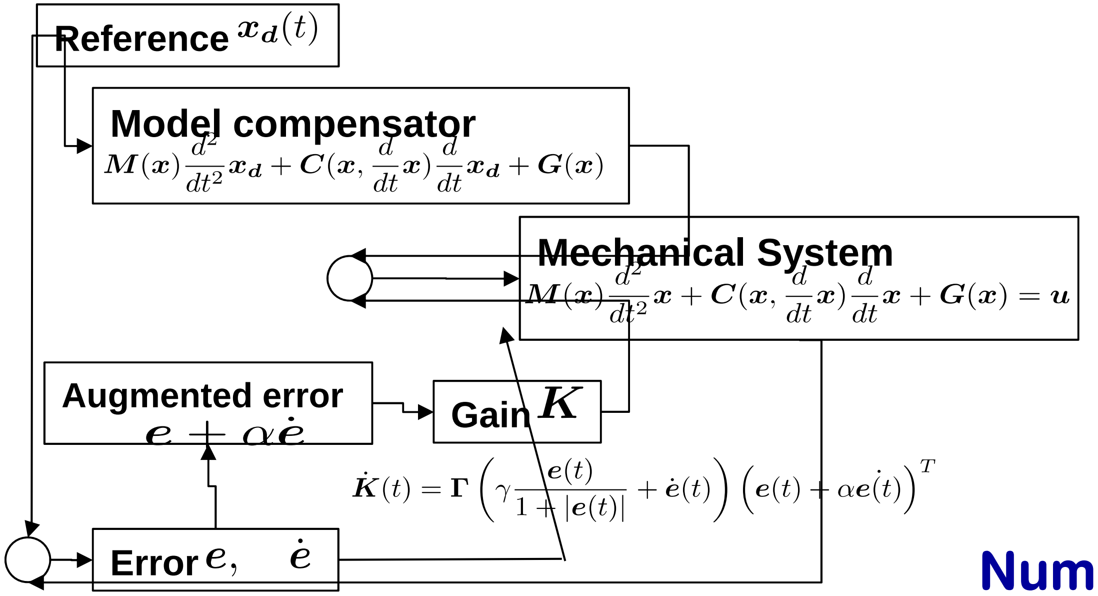


SSL result

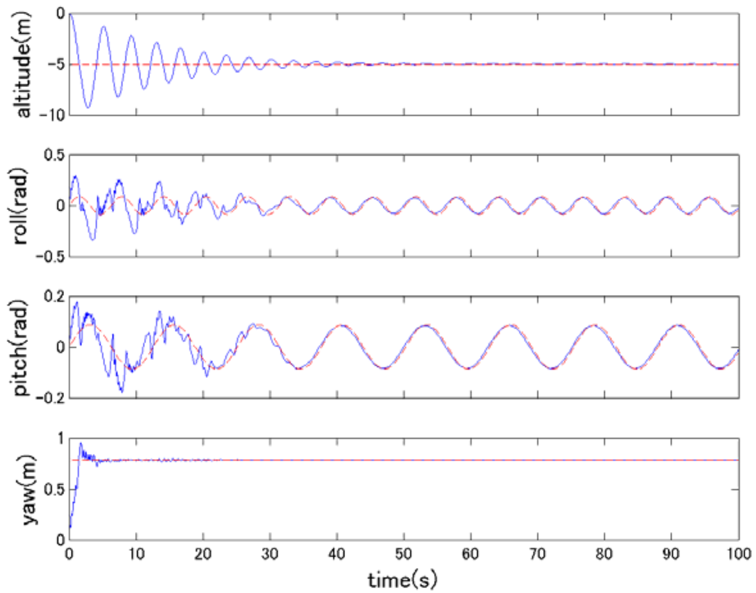
Robust attitude control of UAV

- Accurate control of the platform is necessary in order to keep the attitude of the microphone array stable.
- Uncertainty of the dynamics deteriorates flying performance.

→ Simple Adaptive Control for Quadrotor helicopter



Block diagram of the proposed controller

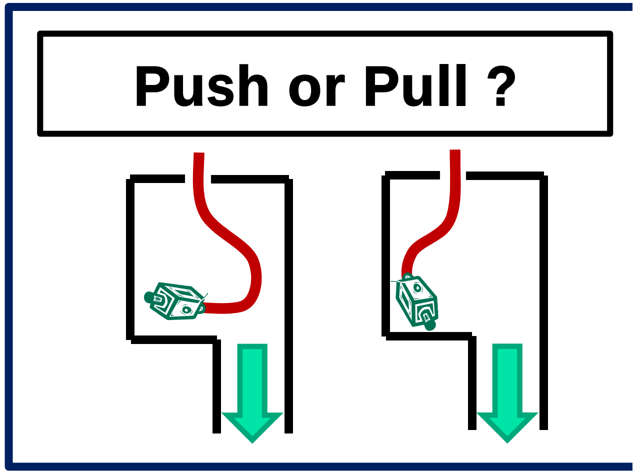


Numerical simulation result

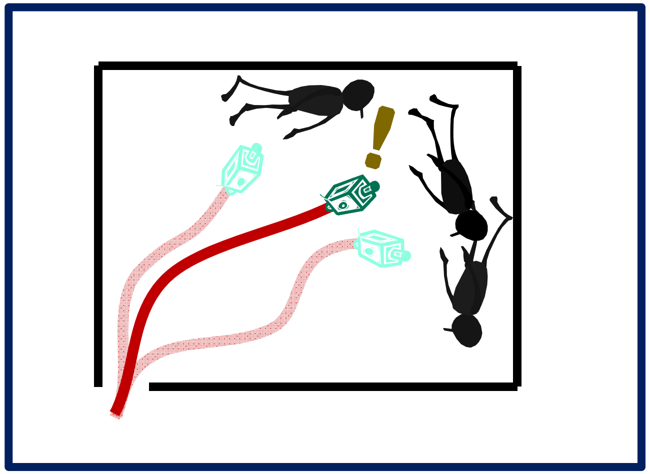
Inertia of the plant is four times larger than that of nominal value. Roll and pitch angles are commanded to follow sine curves

Hose-shaped robots need sounds.

1. **Improve Poster Estimation**
   Posture estimation with inertial sensor or GPS sensor is not robust. Microphone array on the hose may help. Localize microphones' position by sound.



2. **Localize victims**
   Microphone positions provided by posture estimation can be used to improve sound source localization, sound source separation and separated sound recognition.

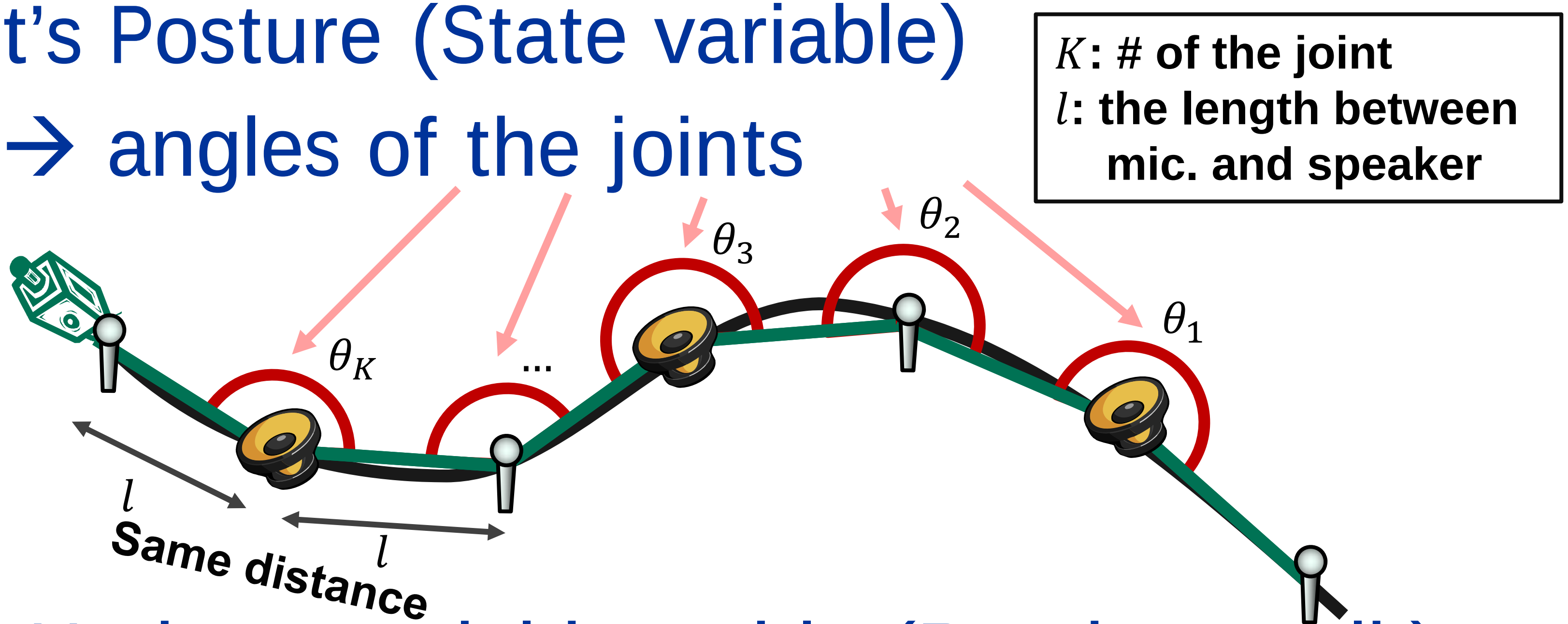


Model of the robot posture

Model with Piecewise Linear Curve

Robot's Posture (State variable)

→ angles of the joints



State Update model is stable (Random walk)

Positions of the microphones and loudspeakers are calculated from angles of the joints and joint length l

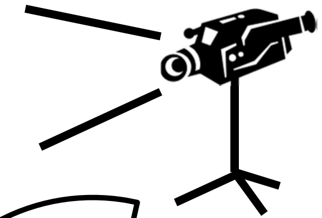
Frog calling

1. **Hard to find or see**
   Small nocturnal animals such as bats, crickets, frogs,
2. **Species identification**
   Needs sound source localization and separation
3. **Song identification**
   Should work on distorted signals due to separation.
4. **Integration of Microscopic and Macroscopic observation**
   ***Microscopic** activities by species and song identification through sound source separation and **macroscopic** activities by "**Firefly**" sound-to-light conversion device.*

Calling Behavior of Many Frogs

- Size of one rice field is about 10m × 20m.
- There are about 10 Japanese tree frogs in one rice field.
- At night, we have to detect the positions and call timings.

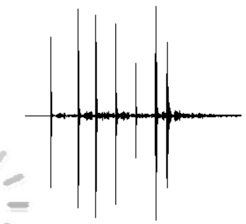
Firefly at a paddy field



We can see where the sound is.
Two frogs are calling alternately.

Firefly2
Motivation

- **Problem:**
  Firefly1 visualizes any sounds without distinguishing calls.
  ➔ multiple species chorus at the same time!
- **Two frogs chorus from Apr to Jun @ Iwakura&Oki**



R. schlegelii
(Schlegel's green tree frog)

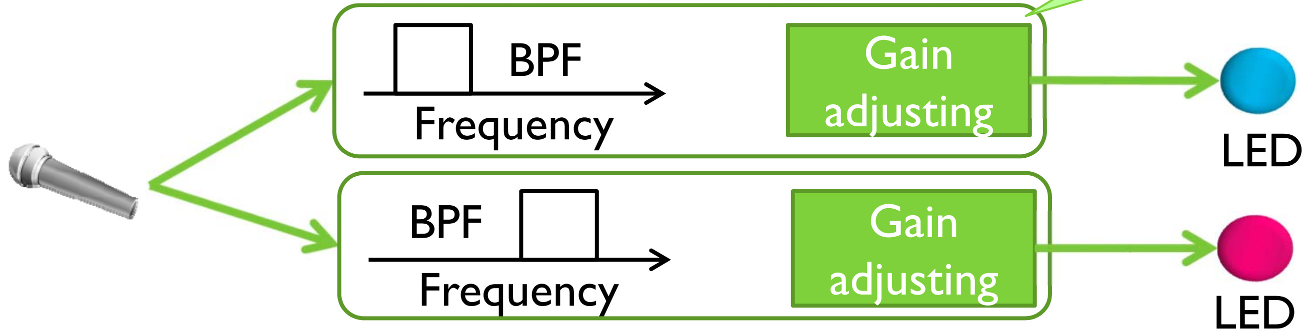
H. japonica
(Japanese tree frog)

- **Can we visualize each chorus separately?**

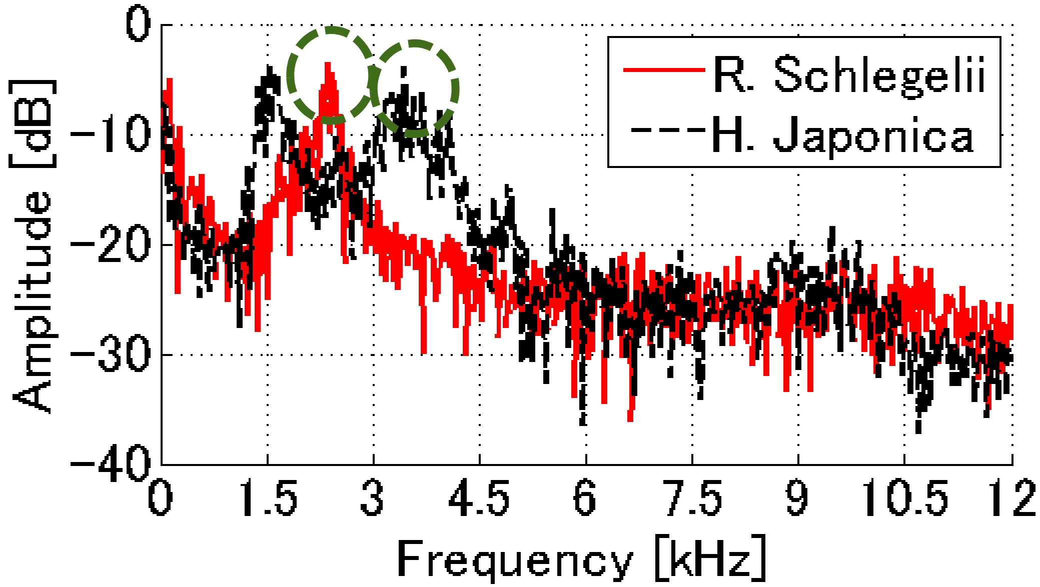
Firefly2
Key idea

Implemented on PSoC microcomputer by H.Awano

- Add **band-path filters** for each species



- Cut-off frequencies are decided from solo-calls recorded by indoor experiments.



Firefly2 Demonstration

- **Indoor experiment** (5 May 2012)
  We took Two R. shlegelii and two H. japonica at Iwakura
  Firefly2 are placed in front of each frogs

- H. japonica: Red LED (They called a lot.

- R. shlegelii: Green LED (Only two times)

Firefly was cited by 60th Ann. Essay

ARTICLE IN PRESS

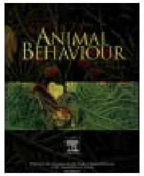
Animal Behaviour xxx (2012) 1–14

Contents lists available at SciVerse ScienceDirect

Animal Behaviour

journal homepage: www.elsevier.com/locate/anbehav

ELSEVIER



60 YEARS OF ANIMAL BEHAVIOUR

Anniversary Essay

All's well that begins Wells: celebrating 60 years of *Animal Behaviour* and 36 years of research on anuran social behaviour

Mark A. Bee[a,*], Joshua J. Schwartz[b], Kyle Summers[c]

[a] *Department of Ecology, Evolution and Behavior, University of Minnesota, St Paul, MN, U.S.A.*
[b] *Department of Biology, Pace University, Pleasantville, NY, U.S.A.*
[c] *Department of Biology, East Carolina University, Greenville, NC, U.S.A.*

Prospectus

Frog choruses truly are dynamic environments for social communication, and we suggest the following directions for future research on these wonders of nature. First and foremost, we advocate the continued development and deployment of new technology to study frog choruses. Until recently, efforts at understanding chorus interactions have been limited to recordings of interactions occurring over relatively small spatial scales involving just a few individuals (e.g. dyadic or triadic interactions among neighbours). Recording interactions over large spatial (and also temporal) scales was too technologically challenging, labour intensive, or both. New technological advances promise to change all this by enabling researchers to explore the complexity of chorus organization in ways only imagined in the late 1970s (Schwartz 2001; Jones & Ratnam 2009; Bates et al. 2010; Mizumoto et al. 2011). Particularly important in this regard are new microphone arrays that not only localize calling males in a chorus, but also recover their original signals for subsequent acoustical analyses (Jones & Ratnam 2009). This is no small technical feat! Future studies should exploit this remarkable new technology to understand better how frog choruses function in the contexts of communication networks (Grafe 2005; Phelps et al. 2007) or social networks (Krause et al. 2009) of signalling males and how females navigate these complex networks when selecting mates. Studies using multichannel recordings and monitoring would enable us for the first time to assess the spatial extent of fine-scale call-timing interactions and their dynamics (i.e. if and how they change) during prolonged periods of chorusing.

Bird songs

1. **Hard to find**
   They usually hide.
2. **Species identification**
   Needs sound source localization and separation
3. **Song identification**
   Should work on distorted signals due to separation.
4. **Song activity detection (Which species sing where, when, why and How)**
   Continuous observation with ***distributed microphone array systems*** *is needed.*
5. **Understand grammar and meaning of bird song**

HARK is used to capture bird songs

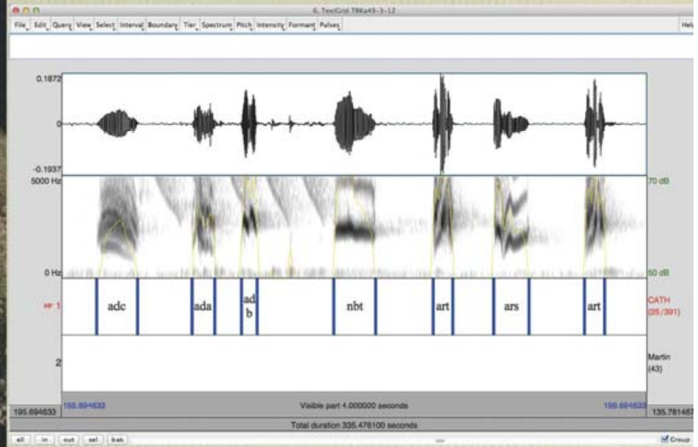
Bird song Activity Detection

Why Bird Song Activity Detection is needed?

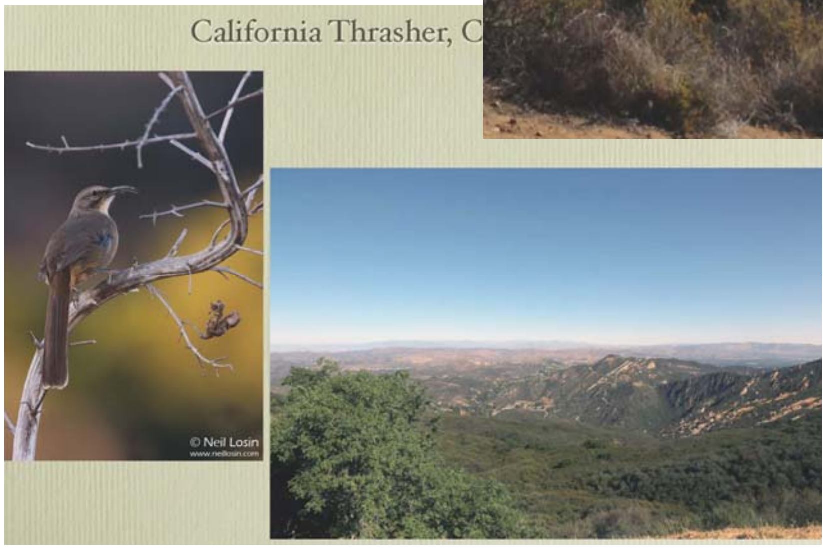
Aim is to understand the grammar and meaning of bird song.

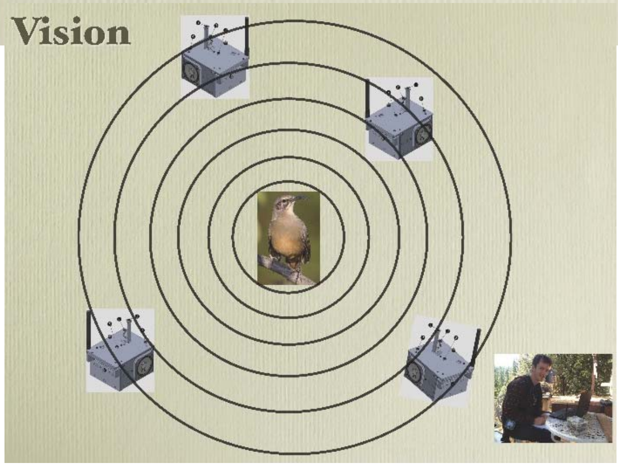
- Approach 1 -- Speak to bird in English
- Approach 2 - Learn bird language
  - "words" for food and various predators
  - "names for offspring" (green-rumped parrotlet)
  - Learn the "words" and the "grammar"

Annotated CATH song



Currently time intensive and needs to be automated





Take-Home Messages

We are engaged in extending *HARK* robot audition software to for outdoor CASA (Computational Auditory Scene Analysis) so that robots can be deployed to real-world to help people recognize and understand auditory scene in natural and disastrous environments.

All's Well That Ends Well