Robot audition and its deployment

Kazuhiro Nakadai

Principal Researcher, Honda Research Institute Japan Co. Ltd.
Visiting Professor, Tokyo Institute of Technology
Visiting Professor, Waseda University

2nd Workshop on Alternative Sensing for Robot Perception: Beyond Laser and Vision
Outline

1. Background of Robot Audition
2. Introduction to Robot Audition Research
3. Open Source Software for Robot Audition HARK
4. Deployment of Robot Audition
5. Summary
Background

Humanoid robot

- Interaction with human is expected to be a partner.

Robot as our partner

Service, Interaction, Information, Entertainment…

- House keeping
- News provider
- Welfare
- Company

Necessity of auditory processing → robot audition
When a robot listens to sound with its ears, it should deal with the mixture of sounds.
Robot Audition

- Proposed by Prof. Okuno (Kyoto Univ. → Waseda Univ.) and Nakadai at AAAI-2000
  - [http://winne.kuis.kyoto-u.ac.jp/SIG/](http://winne.kuis.kyoto-u.ac.jp/SIG/)

- A research field bridging Robotics, AI and Signal processing

- Continuously expanding
  - **Japan**: Kyoto Univ., Honda RI, Tokyo Tech., ATR, AIST, Kumamoto Univ., Waseda Univ., etc
  - **Europe**: CNRS-LAAS (France), INRIA (France), Univ. of Erlangen-Nuremberg (Germany), Ruhr-Universität Bochum (Germany), ITU (Turkey), Imperial College London (UK), etc
  - **North America**: Sherbrooke Univ. (Canada), MERL (USA), Virginia Tech. (USA), Willow Garage (USA), etc
  - **Oceania**: UTS (Australia)
Our Activities for Robot Audition

* Since 2014, robot audition is registered as an official keyword in IEEE-RAS.

Special Session on IEEE Int’l Conf. on Acoustics, Speech and Signal Processing (ICASSP 2009) @ Taipei, Taiwan (ICASSP 2015) @ Brisbane, Australia

HARK Tutorial (OSS)
Korea: 2008
Japan: once a year since 2008

• Migration to Taxai at Willow Garage 2010 @ Palo Alto, USA
• International workshop on Music Robot 2010 @ Taipei, Taiwan
1. Background of Robot Audition

2. Introduction to Robot Audition Research

3. Open Source Software for Robot Audition HARK

4. Deployment of Robot Audition

5. Summary
Robot is surrounded by various noises.

Target Speech

Ego-noise such as motion and voice) (near field, loud)

Reverberation (echo)

Directional noise

Diffuse noise (BGN, omni-directional)

Different characteristics → one-by-one approach

• **Sound Source Separation** mainly for directional noise
• **Dereverberation**
• **Ego-noise suppression**
Sound Source Separation

Separation process

\[ y(\omega) = W(\omega)x(\omega) \]

Incremental SSS:

Update \( W \) to reduce mixing cost \( J(W) \)

\[ W_{t+1} = W_t + \mu J'(W) \]

\( \mu \) : step-size parameter
Sound Source Separation with Adaptive Step-size control

Fixed step-size: Difficult to adapt to environmental changes like robot motions and moving sources => GHDSS-AS

[IEEE-TSLP Nakajima 10]

Manually-tuned

Small value

Large value

Adaptively-controlled μ

Adaptive Step-size (AS) Newton’s method

Recorded sound

SSS

o~te~ra~i
d~o~u
Reverberant conference room (RT > 1s), around 20m x 10m.

Recorded

http://www.youtube.com/watch?v=xpjPun7Owxg
**Ego-noise suppression**

Honda Research Institute

Robot’s voice & motion noise
- closer to mics
- Higher power

Key idea
Robot knows what it utters and what kind of motions it does.

Interactive Dancing Robot

Semi-blind ICA ⇒ barge-in-able robot

Template-based ego-motion noise suppression

### Known signal (utterance)

\[
\begin{align*}
Y(\omega, f) &= \begin{pmatrix} A & H(\omega,0) & \cdots & H(\omega,M) & N(\omega,f) \\ S(\omega, f) & 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ S(\omega, f-M) & 0 & 0 & \cdots & 1 \\ \end{pmatrix} \\
\end{align*}
\]

\[\text{Noise signal} \]

\[\text{Known signal} \]
Mismatch between two blocks

- Noise suppression
- Automatic speech recognition (ASR)

Missing Feature Theory (MFT) for better integration
One of the most important issues is automatic MFM generation.
An example of automatic generated MFM

- Captured spectrogram
  - Left: Arayuru Genjitsu wo ...
  - Center: Isshukan bakari ...
  - Right: Terebi gemu ya pasokon de ...

- MFM
  - 1 (reliable)
  - 0 (unreliable)

- Speech
  - Pass

- Leakage
  - Masked
Outline

1. Background of Robot Audition
2. Introduction to Robot Audition Research
3. Open Source Software for Robot Audition *HARK*
4. Deployment of Robot Audition
5. Summary
Open Source Robot Audition Software HARK

HRI-JP Audition for Robots with Kyoto University

HARK
Honda Research Institute Japan Audition for Robots with Kyoto University

http://www.hark.jp/

hark = listen (old English)
Research: Free (Commercial: Licensing)

Array
Sound Source Localization
Sound Source Separation
Automatic Speech Recognition
Dialog

Developing under collaboration between Kyoto Univ., HRI-JP, and Tokyo Tech.
HARK History and Tutorials

1. Apr. 2008, First release (0.1.7)
   – 1st Tutorial: Nov. 17th, 2008, Kyoto University, Kyoto, Japan
   – 2nd Tutorial: Dec. 5th, 2008, KIST, Seoul, Korea

2. Nov. 2009, 1.0.0 Pre-release
   – 3rd Tutorial: Nov. 20th, 2009, Keio University, Yokohama, Japan
   – 4th Tutorial: Dec. 5th, 2009, Univ. de Pierre et Marie Curie, Paris, France

3. Nov. 2010, Major version-up (1.0.0) – performance, rich documents
   – 5th Tutorial: Nov. 20th, 2010, Kyoto University, Kyoto, Japan

4. Feb. 2012, Version-up (1.1) – performance, 64bit processing, ROS
   – 6th Tutorial: Feb. 29th, 2012, Univ. de Pierre et Marie Curie, Paris, France
   – 7th Tutorial: Mar. 9th, 2012, Nagoya University, Nagoya, Japan

5. Mar. 2013, Version-up (1.7) – Window, Kinect, PSEye
   – 8th Tutorial: Mar. 19th, 2013, Kyoto University, Kyoto, Japan

6. Oct. 2013, Major Version-up (2.0) – HARKDesigner, Microcone
   – 9th Tutorial: Oct. 2nd, 2013, LAAS-CNRS, Toulouse, France
   – 10th Tutorial: Dec. 5th, 2013, Waseda University, Tokyo, Japan

7. Nov. 2014, Version-up (2.1)
   – 11th Tutorial: Nov. 21th, 2014, Waseda University, Tokyo, Japan

8. Nov., 2015 Version-up (2.2) planned
Features in HARK (1)

- **GUI programming environment (HARK Designer)**
  - Web-based programming environment
    (jQuery, node.js, HTML5)
  - **Chrome/Safari/Firefox** on Linux/Windows/Mac
  - Small overhead in module communication (frame-based processing)
    provided by FlowDesigner [Cote04]

An example of robot audition system with HARK

a) Module network

b) Property setting
Features in HARK (2)

- **Support many multi-channel sound input devices**

- **Advanced signal processing technologies**
  - Localization: GEVD/GSVD [Nakamura’11], 3D localization
  - Separation: GHDSS [Nakajima ‘09], HRLE [Nakajima ‘10], etc.

- **Easy to install**
  - Just use a package management tool "apt-get"!

- **Rich documentation**
  - Manual and cookbook over 300 pages in Japanese and English

- **Packages**: ROS, OpenCV, Python, …
1. Background of Robot Audition

2. Introduction to Robot Audition Research

3. Open Source Software for Robot Audition, HARK

4. Deployment of Robot Audition

5. Summary
Human-Robot Interaction according to musical beats
• Adaptive beat tracking
• HRP2, Nao : Thereminist
Issues in mic array processing

- **Given microphone positions**
  Measurements are required.

- **Synchronous recording**
  Special multichannel A/D is necessary.

EKF-SLAM based online mic-array calibration
Quadrotor with 16 mics

SSL with iGSVD-MUSIC

Highly noisy sound sources (-15dB) can be localized.
Sound Source Localization and Visualization

- 2D sound source localization using 1D sound source localization, motion, and gyro information
Robot Audition based IVI system [IROS’15]

Robot-Audition-Based Human-Machine Interface for a Car

K. Nakadai, T. Mizumoto, and K. Nakamura
Honda Research Institute Japan Co., Ltd.

• Talk button less, highly noise-robust voice recognition,
• multi-party dialog with a robot agent
• Hybrid of local and cloud services
Spatio-Temporal Analysis of Frog Chorus

Firefly

Discovery of Three-Group Chorus (=Tri-phase synchronization)
Introduction an overview of robot audition

Introduce open source software for robot audition HARK

Introduced deployment of robot audition technologies to robotics and other fields.
Audio processing is powerful as well as visual sensors, and it is essential to HRI and HMI. Robot audition is a research field to consider techniques working in various real-world scenes. When you are interested in robot audition, please join us (we continuously have sessions in IROS), and try to use HARK.

http://www.hark.jp/