Introduction to Open Source Robot Audition Software “HARK”

Kazuhiro Nakadai\textsuperscript{1,2}, Toru Takahashi\textsuperscript{3}, Takeshi Mizumoto\textsuperscript{3}, Takuma Otsuka\textsuperscript{3}, Hiroshi G. Okuno\textsuperscript{3}, Keisuke Nakamura\textsuperscript{1}, Takami Yoshida\textsuperscript{2}, Gökhan Ince\textsuperscript{1}

\textsuperscript{1} Honda Research Institute Japan Co., Ltd.  
\textsuperscript{2} Tokyo Institute of Technology  
\textsuperscript{3} Kyoto University

Sep. 8, 2011 RSJ annual conf.
Robot Audition [AAAI 00]

- Not a headset microphone, but *robot’s own ears!*
  - Noise-robustness
    - Ego-noise (actuators, self-voice)
    - Environmental sounds
    - Simultaneous speech (barge-in)
  - Cocktail Party Robot
  - Prince Shotoku Robot

- Towards Auditory Scene Analysis
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

  hark = listen in old English

  Research purpose: Free (Commercial: Licensing)

- Apr., 2008 First release
  - [http://winnie.kuis.kyoto-u.ac.jp/HARK](http://winnie.kuis.kyoto-u.ac.jp/HARK)
  - Tutorials in Japan, Korea, France (Humanoids’09)

- Nov., 2010 Major version up to **1.0.0**
  - >50 modules
  - **Linux** (officially support Ubuntu 10.04 and higher)
The following functions are provided by using a robot-embedded microphone array even in a highly-noisy environment such as simultaneous speeches:

- Sound Source Localization (SSL)
- Sound Source Separation (SSS)
- Automatic Speech Recognition of each separated speech
Features in HARK (1)

- Modular architecture based on Flowdesigner [Cote 04]
  - GUI programming environment (modules written in C++)
  - Suitable for frame-based processing like audio and vision
  - No overhead in module communication

- Support many multi-channel sound input devices
  - ALSA based sound devices
  - TED TD-USB devices
  - SiF RASP series
  * Can use any layout and any number of microphones

Example of robot audition system with HARK
Features in HARK (2)

- Advanced signal processing technologies which take dynamic environments into account
  - MUSIC, GHDSS, HRLE, MFT-ASR etc.
- Easy to install
  - Just use conventional package management tool “apt-get”!
- Rich documentation
  - Manual and cookbook over 300 pages in Japanese and English
- High interoperability with robot middleware
  - HARK-ROS: seamless integration of HARK and ROS
  - HARK-MUSIC: music related functions like beat tracking
  - HARK-Binaural: binaural sound localization
  - Wrapper for OpenRTM (release is under consideration)
  - Developing Windows version of HARK (possibly in this year)
Referee for Rock-Paper-Scissors Sound Game

Copyright (C) 2011 Honda Research Institute Japan Co., Ltd. All Rights Reserved.
Four Simultaneous Meal Order Taking

Four Simultaneous Speech Recognition

(Meal Order Taking Task)
Experiment with Texai

- Reverberant conference room (RT > 1s), around 20m x 10m.

[Diagram showing direction and time for Talker 1, Talker 2, Talker 3, and Talker 4, with a mention of Garbage recorded in a reverberant environment.]

http://www.youtube.com/watch?v=xpjPun7Owxg
Visualization of Auditory Scene

Sound archive and reconstruction

Reconstruction using sound location and recognition result

Scene

Reconstruction of sound with specific directions interactively
Sound Lifelog: Visualization for Sound Archives
Towards Auditory Scene Analysis (ongoing work)

- Sound source localization with Generalized EigenValue Decomposition (GEVD)
- Sound source identification with Hierarchical GMM
Summary

• Introduced open source robot audition software HARK
  – Can build a highly noise-robust real-time system using microphone array processing
  – GUI-programming and customization
  – Rich documentation

  – Contribution to robotics and other research fields
  – Just download and use it.

  “Using is believing !”
Acknowledgement

• Special thanks to
  – HARK team (Okuno Lab., Kyoto Univ. and HRI-JP)
  – Dr. Shunichi Yamamoto, Honda R&D
  – Dr. Jean-Marc Valin, CSIRO

• For more information on “Robot Audition”,
  
  http://winnie.kuis.kyoto-u.ac.jp/HARK/
  http://winnie.kuis.kyoto-u.ac.jp/SIG/