Third Symposium on Binaural Active Audition for Humanoid Robots (BINAAHR)
September 30, October 1-2 2013
LAAS-CNRS (Europe Room), Toulouse, France
http://projects.laas.fr/BINAAHR

Sponsored by
- French Research Group (GdR) on Robotics
- Région Midi-Pyrénées
- Université Toulouse III - Paul Sabatier
- Laboratory for Analysis and Architecture of Systems (LAAS-CNRS)

Organized in Cooperation with
- Institute for Intelligent Systems and Robotics, Paris, France (ISIR)

Binaural audition, i.e. hearing capability with a pair of ears in human and animals as well as with a pair of microphones for machines, is a rudimental function for perception and communication. To make robots, either physical or virtual, be in symbiosis with people, they must be endowed with the ability to localize, separate and process sounds under noisy environments or from a mixture of sounds. This symposium focuses on theoretical and pragmatical aspects of binaural audition. It is organized in the framework of the French-Japanese BINAAHR (BINaural Active Audition for Humanoid Robots) project, jointly funded by the French National Research Agency (ANR) and by the Japan Science and Technology Agency (JST). A two-day series of papers will be presented by three invited external lecturers and by members of the BINAAHR consortium. A full-day tutorial on the open-source robot audition ``HARK´´ library (http://winnie.kuis.kyoto-u.ac.jp/HARK/) will conclude the event.

Machine listening, or binaural active audition are first considered in the following manner:
- Generic design of binaural sensors
- Active binaural sound source localization
- Voice detection
- Binaural speaker recognition
- Ego-noise cancellation
- Musical Interaction

The symposium secondly focuses on theoretical approaches to human binaural perception and on their outcomes in robotics. Theoretical contributions from cognitive psychology which put forward the active and exploratory nature of perception are outlined, as well as some artificial perception strategies which hypothesize the interweave of perception and action. The addressed topics are:
- Psychology of Perception in human hearing
- Geometric space perception based on the Sensorimotor Contingencies theory
- Audio-visual integration and Redundancy reduction based on Psychology of Perception based approaches

The Scientific Committee warmly thanks the three plenary speakers, who kindly accepted our invitation.
(In alphabetic order:)
- Prof. **Jens Blauert**, Ruhr University Bochum, Germany
- Prof. **Radu Horaud**, INRIA Rhone-Alpes, Grenoble, France
- Prof. **Shihab Shamma**, Univ of Maryland, College Park, MD

Important dates
For organization issues, participants should register before **September 20, 2013**. In order to encourage scientific exchanges in the field of binaural audition and in connected areas, the registration is **free of charge**.

Scientific Committee
- Sylvain Argentieri (Univ Pierre and Marie Curie & ISIR, Paris)
- Patrick Danès (Univ Paul Sabatier & LAAS-CNRS, Toulouse)
- Alain de Cheveigné (École Normale Supérieure & LPP, Paris)
- Makoto Kumon (Kumamoto University, Kumamoto, Japan)
- Kazuhiro Nakadai (Honda RI & Tokyo Tech, Tokyo, Japan)
- Hiroshi G. Okuno (Kyoto University, Kyoto, Japan)

Local Organization Committee
- Patrick Danès (Univ Paul Sabatier & LAAS-CNRS)
- Alban Portello (Univ Paul Sabatier & LAAS-CNRS)
- Jérôme Manhès (LAAS-CNRS)
- Brigitte Ducrocq (LAAS-CNRS)
- Marie-Agnès Bellières (LAAS-CNRS)

Registration: binaahr2013-reg@laas.fr
Access to LAAS: http://www.laas.fr/2-27920-Practical-Info.php
Hotels (with special discounts)**:** http://www.laas.fr/2-31712-Hotels.php

Third Symposium on Binaural Active Audition for Humanoid Robots (BINAAHR)
September 30, October 1-2 2013
LAAS-CNRS (Europe Room), Toulouse, France
http://projects.laas.fr/BINAAHR

Technical Program

Day 1 : September 30

09:00-10:15 Introduction
09:00-09:30 Welcome – Introduction to the Symposium
09:30-10:15 Overview of BINAAHR
　　　　　　(*Patrick Danès*, LAAS-CNRS & Toulouse Univ Paul Sabatier, France *and Hiroshi G. Okuno*, Kyoto Univ, Japan)

10:15-11:15 Invited Talk: Psychoacoustic Fundamentals and Models of Binaural Listening
　　　　　　by Prof. Jens Blauert, Ruhr University Bochum, Germany

***11:30-12:30 Low-level Auditory* Functions**
11:30-12:00 Active Strategies to Binaural Localization (*Alban Portello*, LAAS-CNRS, France)
12:00-12:30 Binaural Sound Localization and Tracking for Unknown Time-Varying Number of Speakers (Eui-Hyun Kim, Kyoto Univ, Japan)

12:30-14:00 Lunch

14:00-15:00 Invited Talk: Sound Source Separation and Localization on the Binaural Manifold
　　　　　　by Prof. Radu Horaud, INRIA Rhone-Alpes, Grenoble, France

15:15-16:15 Low-level Auditory Functions (cont'd)
15:15-15:45 Extracting Unknown Number of Sources with Two Microphones (*Takuma Otsuka*, Kyoto Univ, Japan)
15:45-16:15 Learning the Sound Source Location in a Binaural Context (*Karim Youssef*, ISIR-UPMC, France)

16:15-17:15 Higher-Level Interaction
16:15-16:45 Audio-Visual Integration for Robots (*Kazuhiro Nakadai*, HRI-JP/Tokyo Tech, Japan)
16:45-17:15 Multi-Party Dialogue System with Robots (*Kazunori Komatani*, Nagoya Univ, Japan)

Day 2 : October 1

09:00-09:30 Welcome – Overview of Day 2

09:30-10:30 Invited Talk: Cortical processes for navigating complex acoustic environments
　　　　　　by Prof. Shihab Shamma, Univ of Maryland, College Park, MD

10:45-12:15 Towards Active Perception
10:45-11:15 Sensorimotor Contingencies Theory and Robotics (*J.K. O'Regan,* LPP, France)
11:15-11:45 A Sensorimotor Approach to Space Perception (*Alban Laflaquière*, UPMC, France)
11:45-12:15 Active Hearing for Auditory Space Learning and Source Localization (*Mathieu Bernard*, UPMC, France)

12:15-14:00 Lunch

14:00-15:00 Design of a Binaural Sensor
14:00-14:30 Sensor Placement and Hardware Architecture (*Patrick Danès*, LAAS-CNRS & Toulouse Univ Paul Sabatier, France)
14:30-15:00 Artificial Pinnae (*Makoto Kumon*, Kumamoto Univ, Japan)

15:15-16:15 Additional Topics
15:15-15:45 Statistical Music Signal Processing (*Katsutoshi Itoyama*, Kyoto Univ, Japan)
15:45-16:15 Multi-modal emotion recognition based on SIRE model (*Angelica Lim*, Kyoto Univ, Japan)

16:30-17:00 Demos

Day 3 : October 2

10:00-17:00 9th HARK tutorial

Summary of the invited talks

Day 1 : September 30
10:15-11:15 Invited Talk: Psychoacoustic Fundamentals and Models of Binaural Listening
        by Prof. Jens Blauert, Ruhr University Bochum, Germany

In this contribution, relevant fundamentals of the psychoacoustics of human sound localization and binaural processing are recapitulated, namely, directional hearing with sounds from lateral directions, directional hearing in the median plane, distance perception and inside-the-head locatedness, summing localization, auditory precedence, echoes, the effect of interaural incoherence, binaural signal detection, and the suppression of reverberance and coloration.
These fundamental psychoacoustic facts will then be discussed in view of modern technology, particularly, in terms of models of binaural listening and understanding.
Special reference will be given to the new EU Project TWO!EARS, which attempts to mimic relevant functions of the complete auditory chain from the auditory periphery through brainstem and midbrain up to the auditory cortex, thereby considering an adequate integration of bottom-up and top-down processes. The TWO!EARS system is indeed intended to represent an advanced model of what is now called "active listening".

Day 1 : September 30
14:00-15:00 Invited Talk: Sound Source Separation and Localization on the Binaural Manifold
        by Prof. Radu Horaud, INRIA Rhone-Alpes, Grenoble, France

In this talk we address two problems: modeling the acoustic space generated by a full-spectrum sound source and using this learned model for the localization and separation of multiple sources that simultaneously emit sparse-spectrum sounds. We start by laying the theoretical grounds allowing to introduce the concept of binaural manifold. We study the low-dimensional structure of the high-dimensional interaural spectral data, based on a corpora recorded with a human-like robot head with motor capabilities. Non-linear dimensionality reduction is used to show that these data lie on a two-dimensional (2D) smooth manifold parameterized by the motor states of the perceiver, or equivalently, the sound source directions. We propose a probabilistic regression model specifically designed to deal with high-dimensional data exhibiting piecewise linear structure. We derive a closed-form expectation-maximization (EM) procedure for estimating the model parameters, followed by Bayes inversion for obtaining the full posterior density function of a sound source direction. We extend this solution to deal with missing data and redundancy in real world spectrograms, and hence for 2D localization of natural sound sources such as speech. We further generalize the model to the challenging case of multiple sound sources and we propose a variational EM framework. The associated algorithm yields a Bayesian estimation of the 2D locations and time-frequency masks of all the sources. Comparisons of the proposed approach with several existing methods reveal that the combination of acoustic-space learning with Bayesian inference enables our method to outperform state-of-the-art methods.

Day 2 : October 1
09:30-10:30 Invited Talk: Cortical processes for navigating complex acoustic environments
        by Prof. Shihab Shamma, Univ of Maryland, College Park, MD

Humans and other animals can attend to one of multiple sounds, and follow it selectively over time. The neural underpinnings of this perceptual feat remain mysterious. Some studies have concluded that sounds are heard as separate streams when they originate from diverse locations or if they are so different as to activate well-separated populations of central auditory neurons, and that this process is largely pre-attentive. In this talk, I will argue instead that stream formation depends primarily on temporal coherence between responses that encode various features of a sound source. I shall also discuss algorithms for implementing this process that are inspired by auditory cortical representation and mechanisms. Furthermore, I will also post argue that only when attention is directed towards a particular feature of a sound (e.g., pitch) do all other temporally coherent features of that source (e.g., timbre and location) become bound together as a stream that is segregated from the incoherent features of other sources.

The 9th Tutorial on Open Source Robot Audition Software *HARK*

HRI-JP Audition for Robots with Kyoto University

October 2, 2013

LAAS-CNRS (Europe Room), Toulouse, France

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

Open source robot audition software called HARK (HRI-JP Audition for Robots with Kyoto University) was released in 2008 as a result of more than ten year research activities including organized sessions on robot audition at IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS 2004–2013). It provides a number of modules for supporting many multi-channel sound cards, advanced signal processing such as sound source localization and separation, and automatic speech recognition of separated speech. A user can make a real-time robot audition system by connecting the prepared modules on HARKDesigner, which is a newly-developed multi-platform GUI programming environment based on Javascript and HTML5. HARK has been applied to many robots including HRP-2, Robovie, SIG to show its general applicability. On Oct. 2 2013, HARK 2.0.0-alpha will be released. Its main features are

- Supported devices & platforms Microsoft Kinect, PlayStation Eye, ROS (HARK-ROS), 64 bit Linux, Windows

- Released new GUI programming environment HARKDesigner, new noise-robust sound source localization, tools for generating transfer functions, packages for visualization and logging incorporating python and MATLAB®, HARK-python, and HARK-MATLAB

- Updated rich documentations

- Many bugfixes

■ Time Table

```
10:00-10:10 Introduction
10:10-10:40 Overview of HARK
10:40-11:10 Practice 0: System boot and Kinect setting
11:10-12:00 Practice 1: Sound localization system
12:00-14:00 Lunch
14:00-15:00 Practice 2: Sound source separation and speech recognition system
15:00-15:20 break
15:20-16:20 Practice 3: Dialog system based on HARK
16:20-17:00 Live demonstration
```

■ Requirements for Participants

Participants are requested to bring a ***laptop PC***, and ***Microsoft Kinect***, ***Sony PlayStation Eye***, or ***Dev Audio Microcone***, which are used in Practices 0-3. Laptop PC (Windows 7 or 8) should be equipped with DVD drive (or an external DVD drive), USB, Core2Duo or higher CPU, and > 1GB memory. The following softwares should be installed in advance : Python (version 2.7.x, http://www.python.org/download/), Kinect SDK (only for Kinect users. The version should be the latest. http://www.microsoft.com/en-us/kinectforwindows/develop/developer-downloads.aspx), nodejs (later than version 0.10.10, http://nodejs.org/download/), Google Chrome (later than version 27), and any text editor.

■ Registration (Free of charge)

Email to HARK13Toulouse-Reg_at_zeus.kuis.kyoto-u.ac.jp (replace _at_ with @)

Subject: 9th HARK Tutorial

I will participate in the 9th HARK tutorial.

*Name:

*Affiliation:

*E-mail address:

*PC (type, vender, specification):