Development of Robot Audition from Computational Auditory Scene Analysis

Hiroshi G. Okuno

(Kyoto University, Graduate School of Informatics, Professor)

【Outline of survey】

Robot audition is a capability in which a robot can hear sounds with its own microphones (ears) mounted on its body. This capability is critical in deploying robots in the real world. Robots usually hear a mixture of ounds in the real world. Computational auditory scene analysis (CASA) studies sound source localization, separation, and recognition of separated sounds. We are applying the results of our previous research on CASA to robot audition in order to implement the capability of listening to several things simultaneously, like "Shotoku-Taishi" (Prince Shotoku), who could, Japanese legend says, listen to the petitions of ten people at the same time. We are focusing on the following research topics:

1) Develop CASA primitives that require minimum prior information for individual robots and environments. 2) Exploit hierarchical integration of audio-visual information to localize, separate, and recognize sounds uttered by multiple moving talkers. 3) Develop symbol acquisition and communication capabilities by exploiting the embodied interaction. 4) Develop multi-person interaction capabilities based on the recognition of speech, music, and environmental sounds.

【Expected results】

A higher level of symbiosis between humans and robots will be attained by achieving the following capabilities:

1) Easy installment of a robot audition system so that any kind of robot can hear multiple sounds the way Shotoku-Taishi could.

2) Recognition of various kinds of sounds including speech, music, and environment sounds.

3) Utilization of unused sound information by converting sounds to symbols.

4) Sound enhancement by modeling and cancelling out self-generating sounds caused by motors.

【References by the principal investigator】

- D.F. Rosenthal and H.G. Okuno (eds.): *Computational Auditory Scene Analysis*, p.399+xiii, Lawrence Erlbaum Associates, Mahwah, NJ, USA, 1998.
- H.G. Okuno: Robot audition from the viewpoint of Computational Auditory Scene Analysis, *Journal of Acoustic Society of Japan*, **63**:1(Jan. 2007) 29-34.
- H.G. Okuno, K. Nakadai, *et al*: Human-Robot Non-Verbal Interaction Empowered by Real-Time Auditory and Visual Multiple-Talker Tracking, *Advanced Robotics*, **17**:2(2003) 115-130

【Term of project】 FY2007－2011 　　　　　【Budget allocation】 47,500,000 yen

(2007 direct cost)

【Homepage address】 　　　http://winnie.kuis.kyoto-u.ac.jp/