

Computational Audio/Visual Scene Analysis (CAVSA)

Contribution to projects from the following areas:

- Human Man-Machine Interfaces
- Humanoid robotics

Contribution to the following scientific areas:

- Computational scene analysis, as well auditory (CASA) as visual (CVSA)
- Multimodality (fusion of auditory and visual analysis) (CAVSA)

Relation to the actual call (work programme):

- ICT-2007.2.2: Cognitive Systems, Interaction, Robotics – Intuitive multimodal interfaces

Description of the contribution:

Humanoid interface agents more and more evolve towards being the mediator between the growing complexity of machine-based tasks (like agent-based negotiations, purchase, planning etc.). Whether in the form of a real physical agent or as an avatar, the communication with the user must be as humanoid as possible. A wide range of robotic applications needs scene analysis, and such analysis can be done in the auditory and/or the visual domain. In the majority of implementations, however, technically "optimized" systems (multicamera setups, microphone arrays) are used.

At Saarland University a humanoid robot (an acoustic dummy head mounted onto a pan-tilt-roll unit) is used to model human scene analysis and behaviour. While the auditory scene analysis has already shown remarkable results with only two microphones, further improvements with respect to the applications (source separation and or tracking) can be reached by the fusion of different modalities. The contributions UdS can bring to potential FP7 Call 3 projects are:

- The hardware demonstrator (unimodal at the moment but to be extended to include the visual path during the next months),
- An optimized auditory scene analysis using several cognitive cues and a sophisticated combining of those cues.

During the course of such a project we want to further develop the visual path and the modefusion, so that the final contribution will be a humanoid interface agent being able to separate and track a user and transform his/her spoken and gestual input into commands.

References:

- Schulz, Sylvia; Herfet, Th.: "Humanoid Separation of Speech Sources In Reverberant Environments", 3rd International Symposium on Communications, Control and Signal Processing 2008 (ISCCSP 08), St. Julians, Malta, March 12th-14th, 2008
- Schulz, Sylvia; Herfet, Th.: " Binaural Source Separation in Non-Ideal Reverberant Environments ", Proceedings of 10th International Conference on Digital Audio Effects (DAFx-07), Bordeaux, France, September 10th-15th, 2007

Contact:

Prof. Dr.-Ing. Thorsten Herfet
Telecommunications Lab
Saarland University
Campus C6 3 - 10.02
66123 Saarbruecken

Germany