SEVENTH FRAMEWORK PROGRAMME

BIANNUAL SB PROGRESS REPORT (M13-M18) Publishable Summary

EARS

Embodied Audition for RobotS

Grant agreement number: 609465

Project acronym: EARS

Project title: Embodied Audition for RobotS

Funding scheme: FP7

Date of latest version of Annex I against which the assessment will be made: 24.10.2013

Periodic report: 3rd

Period covered from: 01.01.2015 to 30.06.2015 (M12-M18)

Project's coordinator: Prof. Dr.-Ing. Walter Kellermann, FAU

E-mail: wk@lnt.de

Project website address: http://robot-ears.eu/

EC distribution: Public

Publishable Summary

Objectives

The project Embodied Audition for RobotS (EARS) explores new algorithms for enhancing the auditory capabilities of humanoid robots. The main focus is on advancing the state of the art regarding fundamental functionalities for a natural dialogue between humans and robots in adverse acoustical environments. The development of the needed algorithms and the dissemination and exploitation of the project results shall be accomplished by the following six workpackages:

WP 1 - Embodied Acoustic Sensing for Real-world Environments

WP 2 - Acoustic Scene Analysis

WP 3 - Audio-Visual Disambiguation

WP 4 - Robot Embodied Cognition and Interaction

WP 5 - System Integration and Validation

WP 6 - Dissemination and Exploitation.

The tackled tasks and achievements in the reporting period (M13-M18) are summarized in the following.

Performed Work and Main Results

The objective of **WP 1** is to develop architectures and algorithms for embodied acoustic sensing for robot audition. Within T1.1, work has been started to evaluate optimum array configurations, which are obtained by maximizing the effective rank as described in the first Progress Report (M1-M12), with respect to the obtained signal enhancement performance of a beamformer applied to the array's microphone signals. The signal enhancement performance is investigated by means of speech recognition scores and signal-based measures. For T1.2, a collaboration between FAU, BGU, and UBER has been started to further investigate the simultaneous use of a head and robomorphic array for source localization in a real-world scenario where the robot is located in a hotel lobby and trying to locate customers. For T1.3, the work related to DOA estimation in moving humanoids initiated previously was expanded by including new robustness analysis and experimental investigation. UBER has investigated optimal robot behaviours and their influence on perception in cooperation with BGU. For T1.4, the EARS Matlab toolbox was extended to handle dynamic environments, which allows users to model sources and receivers in translatory and rotational motion. In addition, the equivalence of different representations for sampled sound fields were studied by BGU, which is useful for the description of a sound field in the context of acoustic scene analysis. For T1.5., INRIA continued to work on the extraction of robust audio features for sound source localization and audio-visual alignment.

The research work of **WP 2** is targeted towards acoustic signal processing relevant to robotic acoustic scene analysis and, in particular, for the example of the Nao robot platform. In the previous period, EARS map objects were developed as interfaces between algorithms across the consortium. These maps contain information such as object positions and uncertainties within the acoustic scene. In this period, novel algorithms were developed that utilise and populate the EARS map objects in order to perform acoustic scene analysis. The work on speaker localization in T2.1 was successfully continued from the previous period for both robomorphic and anthropomorphic arrays. Using the localized speaker directions, novel speaker tracking algorithms were developed for T2.3 in this period. Tracking facilitates the estimation of positions and trajectories of speakers moving within a room throughout time. Furthermore, as robots such as the NAO are not usually equipped with geolocation devices, the tracking algorithms were extended in this period to Simultaneous Localisation and Mapping (SLAM) for the simultaneous estimation of the robot's own position as well as that of the surrounding speakers in T2.1/T2.3. The work on the acoustic front-end microphone array technology was continued, resulting in the implementation of a real-time Blind Source Separation (BSS) demonstrator presented at the EARS review meeting in March 2015. The work on acoustic signal enhancement continues to make good progress as planned. Contributions include the development of a data-independent beamformer for spatial filtering in T2.4 as well as a dereverberation approach for spherical microphone arrays using multichannel equalisation of beamformed channels in T2.7. The work on acoustic echo cancellation in T2.5 will be started in M19. An investigation of a new approach to adaptive interference and noise cancellation was started in T2.6 that constructively exploits reference microphones close to sources of robot ego-noise.

WP3 addresses the issue of how to combine auditory data with visual data in order to disambiguate acoustic information that is noisy or imperfect. This WP concentrates on the following tasks: extraction of visual cues, recognition based on vision only, audio-visual localization and audio-visual recognition. During this period, INRIA finalized the audio-visual clustering method based on a novel weighted-data EM algorithm (T3.2), continued the work on face detection, face-pose estimation and face-pose tracking (T3.1), and started to develop an audio-visual tracking technique being able to detect speech turns among a group of people engaged in a multi-party dialogue involving the robot (T3.2). ALD improved the speech dialogue system whose first implementation was provided as Deliverable D4.1.

In **WP 4**, the interaction of the robot with its environment is tackled taking into account the effects of its sensorimotor action on perception. In particular, UBER investigates the learning of internal models for robot interaction and prediction of sensorimotor states, the creation of an attention system, the development of optimal behaviours for event recognition and localization, a voice dialogue and interaction initiated by the robot. For T4.1, the Self-Organising Maps-based internal models described in the previous progress report have been adopted by UBER for performing forward predictions of sensory outcomes of (intended) motor actions and for executing inverse predictions of motor commands required for achieving a desired sensory state in an experiment on visuomotor coordination. Moreover, UBER is working on extending the internal models and exploration strategies described in the previous progress reports to support the auditory modality. Efforts were made to integrate and adapt existing C++ digital signal processing libraries (in particular, Aquila DSP C++) into the NAOqi framework. For T4.2, UBER ported its own egosphere attentional framework on Aldebaran's Nao in C++ using NAOqi. The behaviour resulting from the activation of the robot egosphere was demonstrated at the EARS Review Meeting in Grenoble on March 2015. For T4.3, UBER investigated different exploration strategies as sensorimotor learning behaviours for the humanoid robot Nao. In particular, UBER studied how a recent concept in developmental robotics, referred to as goal babbling, relates to a visuo-motor coordination task in the humanoid robot Aldebaran Nao that requires coordinated control of two subsystems of motors, namely head and arm motors. In the attention system described in T4.2, efforts have been spent by UBER on improving the feedback system of the robot. Expressive output has been exploited to improve the quality of the interaction. For T4.5, ALD is working on features and their detection which initiate an interaction between user and robot.

The aim of **WP 5** is to evaluate how the scientific and technological innovations, developed by the partners within workpackages WP 1-4, can be integrated into Nao to improve its performance for realistic applications of commercial interest. T5.1 was finished at M6 (according to the workplan). For T5.2, BGU, ALD and FAU worked on Milestone MS4 (microphone array designs for first prototype) which was finished in June 2015 (M18). For T5.3, FAU organized a workshop ('code-camp') at INRIA in Grenoble from February 12th to February 13th 2015 at which EARS co-workers from all partner institutions took part. The goal of this workshop was to provide a solution to directly interface MATLAB with NAO, in order to facilitate the technological transfer from academic partners to the commercial platform NAO. The achieved results have been presented at the Review Meeting in March 2015 in Grenoble. Based on this architecture and using the results and methods of WP1, WP3 and WP4, INRIA started to develop the basis for an EARS demonstrator (first prototype, T5.4) by implementing the following modules: data acquisition with the robot's microphones, sound-source localization in Matlab, face detection running on NAO, a simple audio-visual and motor fusion that aligns the sound-source location with the presence of a speaker face and controls the robot gaze towards the audio-visual event.

The aim of **WP 6** is to maximise the impact of the project in the scientific community, in industry, and the society.

The maintenance and further development of the EARS website by FAU was a major activity within the reporting period as its serves as a main basis for the dissemination activities within EARS. Theses efforts provided the basis for in a significant increase in the number of page views and page visits. Moreover, the

EARS project is now represented at Twitter and Facebook.

A second major activity was the publication of the achieved research results by the various project members at international conferences and workshops as well as within journals.

A third major activity was the preparation and organization of various dissemination events like workshops and special sessions and tutorials at conferences. (All scientific publications and dissemination activities conducted within the EARS project so far are listed in Annex G and Annex H.)

List of Acronyms and Abbreviations

BSS	Blind Source Separation
EARS	Embodied Audition for RobotS
SLAM	Simultaneous Localisation and Mapping