FP7-ICT-2013-C TWO!EARS Project 618075

Document name

Final Project Report

Alexander Raake and Two!Ears Team *

January 6, 2017

* The Two!Ears project (`http://www.twoears.eu`) has received funding from the European Union's Seventh Framework Programme for research, technological development and demonstration under grant agreement no 618075.

Project acronym:	Two!Ears
Project full title:	Reading the world with Two!Ears

Work packages:	All WPs 1–8
	Final Project Report
Version:	1.0

Delivery date:	Jan. 6, 2017
Dissemination level:	Public
Nature:	Report

Editor:	Alexander Raake
Author(s):	Jens Blauert, Jonas Braasch, Guy Brown, Patrick Danès, Torsten Dau, Bruno Gas, Armin Kohlrausch, Dorothea Kolossa, Ning Ma, Tobias May, Klaus Obermayer, Alexander Raake, Sascha Spors, Hagen Wierstorf

Contents

1 Executive summary

In Two!Ears' systematic auditory modelling approach, human listeners are regarded as multi-modal agents that develop their concept of the world by exploratory interaction. The goal of the project has been to develop an intelligent, active computational model of auditory perception and experience in a multi-modal context. Our novel approach is based on a structural link from binaural perception to judgment and action, realised by interleaved signal-driven (bottom-up) and hypothesis-driven (top-down) processing within an innovative expert system architecture. The system achieves object formation based on *Gestalt* principles, meaning assignment, knowledge acquisition and representation, learning, logic-based reasoning and reference-based judgment. More specifically, the system carries out dynamic auditory scene analysis by combining signal- and symbol-based processing in a joint model structure, integrated with proprioceptive and visual percepts. Our system was implemented on a robotic platform, can actively parse its physical environment, orientate itself and move its sensors in a humanoid manner. The system has an open architecture, so that it can easily be modified or extended so as to be used in a wide range of application domains.

A number of significant outputs result from the project, which have been made publicly available to the research community. These include: (1) Two!Ears' best practice reproducible research approach, (2) the Two!Ears database of audio and visual scenes and scene elements, compatible with (3) Two!Ears' Auditory Front End, a model framework for bottom-up auditory signal processing with interfaces to top-down feedback control, (4) the blackboard and expert system and related training pipeline for scene analysis and quality evaluation, and (5) a number of re-implementable robotics components both in hard- and software, including a reproducible head-rotation unit with robotics-typical ROS software for the KEMAR Head-and-Torso-Simulator and instructions for humanoid and smaller-scale affordable robot units that enable lateral displacement. All major developments within Two!Ears are described on or interlinked from the project website `www.twoears.eu`. The Two!Ears consortium successfully held an inter-disciplinary summer school in fall 2015, presented the project and its test bed to the public at the Nuit Européenne des Chercheurs in Toulouse in 2016, and published a number of peer-reviewed conference and journal papers. Two!Ears members have organized different special sessions at national and international conferences between 2013 and 2016. Two!Ears is thought to have significant impact on future development of ICT wherever knowledge and control of aural experience is relevant. It will also benefit research in related areas such as biology, medicine and sensory and cognitive psychology.

2 Project context and objectives

The Two!Ears project was set up in the context of different emerging developments and trends in research fields such as audio signal processing, psychoacoustics and auditory modelling, auditory scene analysis, machine learning, cognitive modelling, robotics, and Quality-of-Experience assessment. On Twitter we have described Two!Ears as: "Study human hearing with robot & 2 ears, incl. scene analysis, cognition & quality." The main objective was to provide a testbed framework for fostering science towards "Reading the world with Two!Ears". The relation to human hearing and auditory modelling is directly reflected in the project name: hearing with *only* two ears or microphones – instead of using a higher number of sensors, as often done for acoustics applications such as beam-forming or sound-field analysis.

A comprehensive model of human auditory and cross-modal experience can be considered as a major scientific breakthrough and enables innovative application opportunities in ICT, robotics, and many other areas. Our novel modelling system provides a powerful tool for R&D, rapidly adaptable to new scientific questions and new applications. To our best knowledge, nothing comparable previously existed, namely, a computational framework for modelling active exploratory listening to assign meaning to auditory scenes. Our computational framework is targeted as the front-end for a great variety of applications: cocktail-party processors to perform auditory-stream segregation in acoustically adverse conditions, hearing-aids, auditory-object identification and tracking, automatic speaker identification and speech understanding, perceptual analysis of auditory environments, intelligent microphones, perceptual audio coding, robot audition, quality-of-experience (QoE) assessment of performance spaces, spatial audio transmission and reproduction systems.

Auditory perception, auditory object identification, and the assignment of meaning are context- and task-specific. By *meaning*, we imply that the computational model is able to experience the properties of the acoustic environment in a way inspired by how a human listener does (for example, whether particular sources are moving or static, familiar or novel, spatially distinct or diffuse). For example, information about the type of source the system is to detect, such as a voice in a mixture of sounds, may modify the pre-processing stages and exploratory behaviour of the system used to segregate the target voice. On the other hand, if the task of the system is to evaluate the quality of an audio reproduction system, more holistic features may be more relevant, and no suppression of "background" sounds may be applied. Such task-specific meaning will determine the feedback being initiated by the upper (cognitive) model layers to drive the lower levels. According to this paradigm, a modelling framework has evolved that enables applications that consider auditory spatial perception, by use of a model that can be configured to be application-aware in itself – hence closing the loop between applied technology on the one hand and perception and cognition on the other. The applications that will be enabled or significantly improved by the new model are of high social and economical relevance. To foster science and innovation, the software system and results of Two!Ears are made publicly available according to a "reproducible research" paradigm.

The project has combined two sets of objectives: (a) The development of an overall system that can serve as test bed for individual modules in terms of overall system performance on specific tasks, and (b) significant scientific advancement on individual aspects of the system, required so as to deliver the targeted overall system functionality.

Two!Ears distinguishes a *development system* and a *deployment system*. The *development system* is a software system that serves as the platform for addressing implementations of individual functions according to the architecture chosen for the project. The *deployment system* is the interfacing of the *development system* with actual robot systems intended for real-time operation. The deployment system was developed in three parallel stages: (i) a virtual implementation connecting auditory modelling in terms of a virtual agent with virtual acoustics and visual simulation; (ii) a Head-And-Torso-Simulator (HATS) endowed with cameras for stereoscopic vision, allowing rotation motion of the "dummy" head to actively explore the environment; (iii) the implementation on a robot platform involving the same basis of software as used in (i) and (ii), and the HATS from (ii). In addition to head-movements, this final system also enables physical translational motion and, hence, full exploration of real-world scenes.

To prove the functionality of the overall system, two proof-of-concept applications were defined: *Dynamic Auditory Scene Analysis* and *Quality of Experience* assessment. Consequently, the systematic evaluation process that has been implemented comprises tasks according to these two proof-of-concept areas.

The different levels of the Two!Ears system each were planned to comprise a set of complementary and/or exchangeable modules, where different modules at one level may be differently suitable for achieving a specific functionality. Considerable scientific advancement at individual levels of the Two!Ears system was needed, to finally enable a system that can be used for modular testing of model components in terms of performance on auditory scene analysis and/or quality evaluation. This goal has been addressed with Two!Ears' innovative architecture with interconnected bottom-up and top-down components.

2.1 Objectives

A set of detailed objectives has been laid out for Two!Ears for its three-year project duration. These comprised to create and maintain a central database of labeled audio-visual scenes for use throughout the entire project, which was required for development, training and evaluation of models, and also represents a significant output in itself. The database was to contain data such as ear-signals, head-related impulse responses, HRIRs, multichannel recordings, multichannel room-impulse responses, MRIRs, still images and related meta data. The database required means of usage and downloading, which were part of the substantial software engineering tasks in Two!Ears. Methods were provided to combine spherical microphone array measurements with HRIRs in order to derive head-tracked ear-signals. The project increasingly built up a high-quality content database, as described in more detail in Chapter3.

Further objectives were related with the bottom-up signal processing that represents different auditory modelling steps. This objective comprises the realisation of signal-driven transformation of listeners' ear signals into multi-dimensional representations – in view of known functions of the subcortical auditory system. The output of this stage consists of several transformed/filtered versions of ear signals plus perception-based descriptors, to serve as input to higher model stages. Further, the generation of a signal-processing architecture was targeted, that allows higher stages to influence peripheral properties via feedback mechanisms and to adjust the processing to the context and with respect to the listeners' intentions when 'reading' their auditory environments.

At the next higher stage, the goal was to implement an expert system using a graphical-model-based blackboard architecture on top of the aforementioned monaural/ binaural models, to operate in multiple-sound-source environments. The system was planned so as to consist of a number of layers connected by feedback pathways. First layer: Auditory scene is pre-segmented into different auditory streams corresponding to sound sources (foreground) and background. Second, 'event-expert' layer: Information about foreground streams were to be used to define auditory events. These were to be semantically interpreted and annotated, creating a first symbolic scene description. Third layer: Symbolic information about multiple auditory events are combined in the graphical model architecture. Sound events will be put into context to disambiguate interpretation and assign 'meaning' to the auditory scene. The blackboard architecture was planned as the component to impose contextual constraints, with the third layer designed for inclusion of knowledge from other modalities (vision and motor-control).

Active listening entails bottom-up data processing as well as top-down mechanisms. To capture them, a framework was set up that can host suitable feedback loops. It has been a further objective in Two!Ears to design an appropriate system architecture for this requirement, investigate meaningful feedback paths, implement them, and finally evaluating them regarding their functionalities. Input from other modalities than the auditory one was to also be considered as a source of feedback information, particularly, position, direction and speed of head-&-torso movements and of identified optical objects (visual input).

To enable a physical implementation of the active listening agent, the set of modules from the aforementioned activities was to be integrated into a physical test bed assisting in the global evaluation of the Two!Ears architecture. The physical system, above referred to as the "deployment system", consists in a number of test beds, ranging from a pan-motion capable binaural head or a HATS system to a visio-auditive robot. The complexity and versatility of each such platform fundamentally depends on its motion and sensing capabilities. Embedded (e.g., FPGA-based) hardware was considered where appropriate, to enable sensor parametrization as well as signal-driven low-level cues computation. Requirements like real-time processing, data time stamping, etc. had to be addressed. In parallel, the Two!Ears framework was implemented as a comprehensive software modular architecture.

The application domains for which the Two!Ears model were to be evaluated have been: (a) *Dynamic auditory-scene analysis* (DASA) and (b) *Quality of Experience* (QoE) assessment. For DASA , the general objective was to evaluate, for a number of 'search and rescue' tasks, how well the framework allows perception- and cognition-based questions about scenes to be answered. Insights achieved here were planned to contribute to the selection of best model components, and were expected to contribute to improve the auditory processing and grouping modules described above. Particular emphasis was put on the contribution of interactive rotatory and translatory head movements for solving scene-analysis problems, based on the mobile robotic system. For the second proof-of-concept application domain (b) QoE, the objectives were to develop appropriate assessment methods for reference-free audio reproduction quality tests, and collect respective labels of scenes presented via such technology. Further, it was planned to apply the integrated Two!Ears modules to QoE assessment, covering multichannel loudspeaker systems driven with Wave Field Synthesis or Higher-order Ambisonics.

Different channels and activities have been undertaken to disseminate the results of Two!Ears to academic institutions and industry, to ensure reaching the impact the project aimed at. Here, based on the links of the involved partners, ongoing external ideas and trends were to be included, to ensure the breakthrough character envisaged at project start. As a part of this activity, the collaboration with project-external research groups was fostered, to ensure training of other groups on the Two!Ears framework and toolboxes. Different high-quality publication and communication channels were identified and addressed by respective dissemination activities.

3 Scientific results

The guiding idea of the Two!Ears project has been to develop an intelligent computational model of auditory perception and experience in a multi-modal context. Hereby the processes of active exploration and interactive analysis of auditory scenes were the main focus. Active exploration was understood in such a way that the Two!Ears system collects information from the environment (i) via top-down adaptation of bottom-up signal processing components as well as higher-level processing stages and (ii) via the physical orientation and displacement of a mobile platform within the environment. The Two!Ears system is available in two complementary versions: As an emulated (virtual) implementation, and as a robotics hardware platform that actively parses its physical environment.

All critical objectives addressed by Two!Ears have successfully been achieved. In this chapter, after a short summary the highlight project results in different areas are outlined, complemented by the impact-related considerations in Chapter 4.

(1) Exploration of auditory scenes using simulated environments with virtual agents and real scenes with robots. The first area addresses the two ways in which the acoustical input signals are provided to the model. First, by the creation of an interactive binaural simulation tool that generates realistic ear signals on the basis of an abstract description of the virtual scene. A large database of acoustic measurements has been collected to allow interactive simulations of a wide range of acoustic environments that can be explored by the model. Secondly, a collaboration with the robotics community has been established by interconnecting the ROS robot ecosystem and the MATLAB dominated world of auditory modelling.

(2) Allowing a unified computation of classical auditory features as simple requests by higher modeling stages. The work in this area represents a cumulation of many years of auditory modelling work, and significantly advances the approach of existing auditory toolboxes (such as the AMToolbox) by providing an integrated framework, and thus an easy interface to access individual features. This enables the use of machine learning techniques at large scale for higher modelling stages, and integrating with the auditory modelling community addressing more real-life scenarios, which is a very relevant current trend in the area of human perception and cognition modelling.

(3) Machine learning approaches for binaural auditory and cognitive modelling State-of-the-art machine learning methods were applied in Two!Ears– particularly deep learning – to computational auditory-scene-analysis and hearing tasks. This has allowed us to leverage highly effective techniques from the field of automatic speech recognition, and also allows for highly relevant new approaches such as combined localisation and identification of several sound sources. This area of achievement also comprises the implementation of the higher-level blackboard architecture, and its integration with the applied machine learning approaches.

(4) Auditory and crossmodal feedback There is strong physiological evidence that the auditory system provides various feedback in the course of auditory signal processing. Actually, top-down connections between almost all stages of auditory processing have been identified. Information from other senses is considered as well and integrated where suitable. Yet, in terms of technological application of binaural signal processing, feedback was only rarely considered – at least at the time that Two!Ears had started. However, it seems that Two!Ears has served as a significant trigger project in this respect, as we now find increasing activity along those lines, in particular as regards robotics and hearing-instruments.

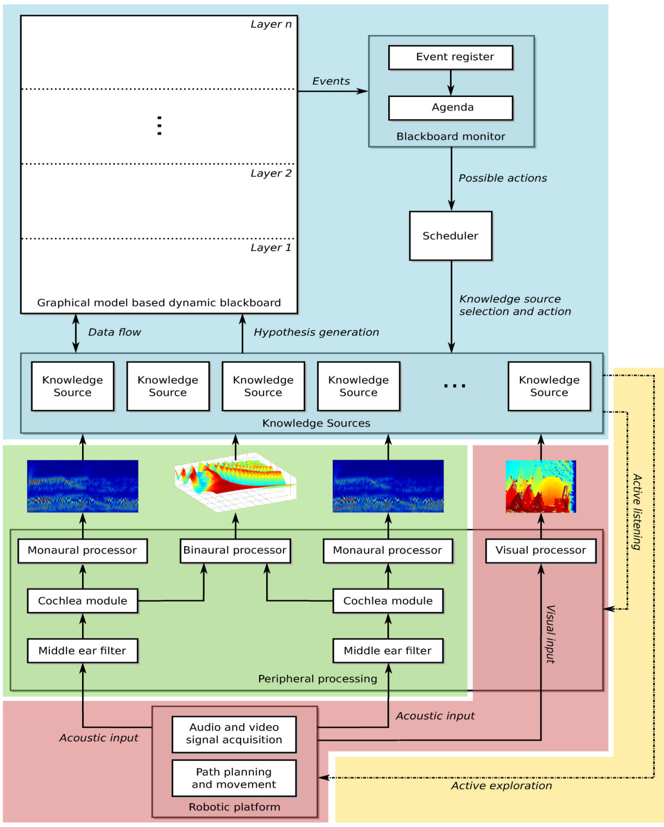

Figure 3.1: Overview of the general system architecture.

3.1 Exploration of auditory scenes using simulated environments with virtual agents and real scenes with robots

To put the Two!Ears model (or any other tentative model of auditory perception and experience) to the test and determine whether it encompasses active exploration and interactive analysis, experiments are essential. Consequently, tools were designed during the course of the project in order to enable a user to conduct realistic offline simulations as well as live tests on physical devices. A best practice was followed, which consists in defining scenarios that demonstrate increasingly elaborate capacities. Such scenarios, in turn, require the virtual (simulated) and physical test beds to be endowed with increasingly complex basic abilities. Within Two!Ears, these can consist in: on-the-fly parameterization of raw binaural acquisition and low-level processing routines; rotational motion of a binaural head; long-range translations of a binaural head; interweaving of perception and motion; incorporation of multi-modality; real time behavior; etc.

3.1.1 Interactive binaural simulator

The acoustic signals at the ears serve as input for the auditory scene analysis performed by the human auditory system. The goal of the Two!Ears project is to develop an advanced, active computational model of auditory perception and experience in a multi-modal context. The model relies mainly on the auditory sense but also considers the visual sense for multimodal integration. The synthesis of ear signals is an important basis for the development and evaluation of the model. The synthesis allows the generation of reproducible conditions in contrast to the input in a more or less controllable real-world scenario. A challenge is the desired exploratory nature of the Two!Ears model. Its interaction with the environment has to be tracked by the simulation environment. This has been solved by implementing a block-based binaural synthesis framework. The ear signals are computed for subsequent ensembles of samples (blocks). Parameters like source position and the position of the listener including its head-orientation can be changed in-between blocks in order to support exploratory actions. Various techniques for the computation of ear signals have been developed and implemented. In particular, the synthesis of ear signals using

1. **pre-recorded binaural signals**
   A straightforward approach is to record ear signals. The recording can either be performed by placing small microphones at a defined position in the ear canal of a listener or by a Head and Torso Simulator (HATS). The synthesis of pre-recorded binaural signals is performed by playing back the recorded signals.
2. **static head-related and binaural room impulse responses**
   Under the assumption of time-invariant linear acoustics, the transfer path from a

sound source to the ears can be characterized by impulse responses. Under free-field conditions these are referred to as Head-Related Impulse Responses (HRIRs) and in reverberant environments as Binaural Room Impulse Responses (BRIRs). The synthesis of ear-signals is performed by convolving the desired source signal with the appropriate left/right HRIR/BRIR.

3. **dynamic binaural room impulse responses**
   The transfer path from a moving sound source to the ears can be characterized by time-variant impulse responses. The identification of such impulse responses requires specific signal processing techniques that cope with the time-variance of the underlying linear system. Time-variant BRIRs can capture the movement/orientation of a sound source on a fixed trajectory for a fixed head-orientation in a reverberant environment. The synthesis of ear-signals is performed by time-variant convolution of the desired source signal with the time-variant BRIRs. Within Two!Ears existing time-variant system identification techniques have been advanced and evaluated Hahn and Spors (2014a,b, 2015, 2016).
4. **data-based binaural synthesis**
   Data-based binaural synthesis allows to compute ear-signals for translated listener positions. It is based on a directional analysis and translation of the reverberant sound field. The sound field captured by a (spherical) microphone array is decomposed into plane waves using (modal) beamforming. With respect to the origin, plane waves can be shifted in space by applying a spatial phase shift. The shifted plane waves are then filtered by the respective HRIRs and summed up for auralization. The result are the BRIRs for a given listener position which can be used for auralization using convolution by a source signal. Data-based binaural synthesis allows to consider small scale translatory movements of the listener around a fixed location. All physical aspects of a reverberant sound field are included. The perceptual properties in terms of localization for various technical parameters have been evaluated by Winter *et al.* (2014).
5. **numerical simulation of acoustic environments**
   The numerical simulation of acoustic environments has been a very active field of research for several decades. Besides the numeric solution of the wave equation, various approximations have been developed. These approximations are based on the assumption that sound propagation can be modeled by rays, which is reasonable if the dimensions of the considered objects are large compared to the wavelength. The numerical simulation of acoustic environments allows to compute BRIRs for a given head-orientation and listener position. These BRIRs can then be used to compute ear-signals by convolution with a desired source signal. The well-known mirror image source model has been implemented.

Each of the methods has its weaknesses and strengths considering the desired acoustic environment, degree of realism and interactivity. Please refer to Deliverable 1.3 for a detailed discussion and comparison. One particular method or a mixture of methods can be chosen on a per simulation basis.

The implementation is based on the *SoundScapeRenderer* (SSR) INT (a), which is a widely applied tool for spatial audio reproduction. For the interactive binaural simulator various enhancements of the SSR have been realised via an MATLAB executable (MEX). The binaural simulation framework for the synthesis of ear signals has been fully integrated into the Two!Ears software framework. It is interfaced with the peripheral processing and feature extraction stage of the Auditory Frond End (AFE) (Section 3.2). The blackboard system developed in Two!Ears (Section 3.3) acts as the core of the integrated framework and provides additional interfaces to include feedback mechanisms as described in Section 3.4 and a robotic interface, as described in Section 3.1.2. The listener position and head orientation is controlled by the robot interface. The integration is described in more detail in Deliverable 3.2. The simulation framework for the synthesis of ear signals has been published under `https://github.com/TWOEARS/binaural-simulator`.

For the synthesis of ear signals, a decent amount of recorded and measured data is required. Furthermore, for evaluation of the model against human performance, perceptual labels are mandatory. A central database has been established for this purpose. The database is seamlessly integrated into the binaural simulation. The database is publicity available under `https://dev.qu.tu-berlin.de/projects/twoears-getdata`.

The simulation framework has been applied successfully in the project. The full integration into all work streams has allowed significant scientific contributions in many tasks. The binaural simulator was used for the development and evaluation of the Two!Ears model. Furthermore, it was used in most of the QoE-related proof-of-concept-scenarios (see `http://docs.twoears.eu/en/latest/database/listening-tests/`).

3.1.2 Comprehensive binaural/multisensory mobile robots

We have assumed that the world of auditory modeling is dominated by MATLAB. Accordingly, the Two!Ears model was implemented in this language. Robotic platforms were designed and completed by a stable real-time software architecture so that they could be interfaced with this model. Their deployment obeyed the following guidelines: silent suitable motion capabilities; use of classical robotics middleware; use of off-the-shelf software only if widely acknowledged; model-driven design of components specific to the project for improved properties; efficient bridge from the robotics software to the Two!Ears model; generic open-source code for reproducible research.

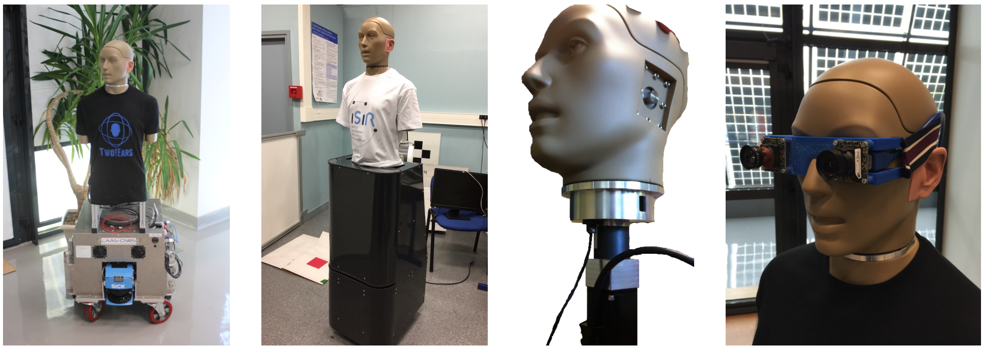

Figure 3.2: From left to right: Two!Ears binaural robots, consisting of a KEMAR HATS and a mobile platform; motorisation of the KEMAR head; stereoscopic setup.

Robot platforms and integrated audio/audiovisual sensors

A controllable degree-of-freedom (DOF) was designed to enable azimuth rotation of the neck of a KEMAR head-and-torso simulator (HATS). This non-intrusive device is screwed on the mounting holes of the KEMAR torso in exactly the same way as its genuine assembly mechanism. Two such motorised HATSs were mounted on distinct differential wheeled robot, to provide omnidirectional motion of the head and long-range navigation. Multimodality was added thanks to the design of 3D-printed glasses from the head CAD model, which support a rig of two micro-cameras with suitable lenses (Fig. 3.2). Instructions to set up a similar motorised binaural head are available at `https://github.com/TWOEARS/kemar-hardware`.

Real time modular software architecture

A *modular software architecture* was installed on each binaural robot. On the top of instrumentation, the *functional layer* consists of C/C ++ components which can run concurrently under severe time and communication constraints. The ubiquitous ROS (Robot Operating System)[1] middleware was selected to ensure their real time control and communication. Typical components of the functional layer are: motion control; audio / visual data streaming and low-level processing; robot navigation. Higher in the architecture, the *decisional layer* hosts primitives working at a more abstract level, under lighter time constraints. It straightly includes the cognitive part of the Two!Ears computational model, as well as top-down hypothesis-driven feedbacks, all written in MATLAB. An intermediate set of abilities was identified in-between, which should be implemented into components of the functional layer for maximum responsiveness,

1 `http://www.ros.org` – ROS runs on top of GNU/Linux.

but can first come in MATLAB if the pace of experiments is slowed down, typically: advanced audio / visual segmentation, detection, tracking, etc. In addition, a specific "bridge" (`https://git.openrobots.org/projects/matlab-genomix/gollum/demo`) had to be inserted between the decisional (MATLAB-based) and functional (ROS-based) layers, so that single / multiple MATLAB processes could control and read data from any set of native ROS /GenoM3 components running on single / multiple CPUs, see Figure 3.3. To improve components re-usability, code robustness and sustainability, and to enable

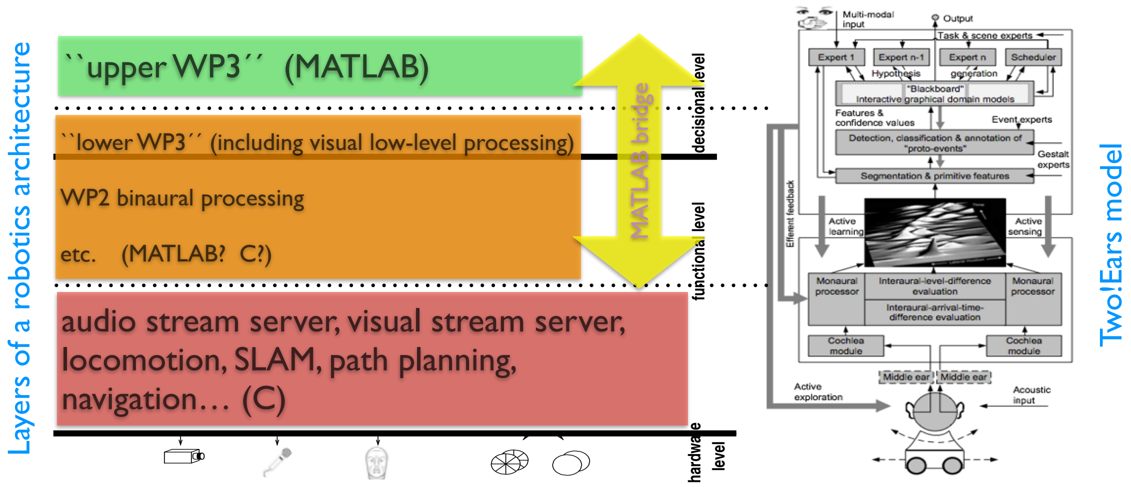

Figure 3.3: From the Two!Ears computational model (right) to a real time robotics software architecture (left).

scalability as well as formal proofs of dependability, Two!Ears-specific components of the functional layer were designed and implemented by means of the GenoM3 (Generator of Modules v3) framework[2]. Importantly, they can be deployed on other middlewares with no additional line of code.

The following GenoM3 components were deployed for Two!Ears: Advanced Linux Sound Architecture (ALSA)-compliant optimized binaural audio streaming (`www.alsa-project.org`, `https://github.com/TWOEARS/audio-stream-server`); significant subset of the MATLAB-based auditory front-end (AFE) (Sect. 3.2); KEMAR neck rotation; robot locomotion, teleoperation, and sensing; active source localisation (interweaving binaural data and motor commands, and entailing sensorimotor feedback); visual detection and tracking of humans. Off-the-shelf ROS components include: stereoscopic data streaming and low-level operations (calibration, rectification, etc.); map

2 `https://git.openrobots.org/projects/genom3/wiki/Wiki` – GenoM3 is one of the core elements of the open-source collection developed at CNRS, and is the result of two decades of research on real-time architectures for autonomous systems.

building; navigation (including localisation, path planning and execution, reactive obstacle avoidance); stereoscopic detection and localisation of objects.

3.1.3 Conclusion

A binaural simulation has been realised which features the interactive computation of realistic ear signals for acoustically complex scenarios. It is accompanied by a rich database of acoustic measurements and perceptual labels. Both provided the grounds for a systematic development of the Two!Ears model and its reproducible evaluation. Such an integrated simulation framework was not publicly available before, and hence is a significant contribution by Two!Ears. It is of general use in psychoacoustic research, for instance for the generation of binaural stimuli for listening experiments.

Besides, two binaural robots were designed and deployed, on the basis of similar motorised KEMAR HATSs and distinct differential wheeled mobile platforms. Vision was incorporated, thanks to an embedded anthropomorphic stereoscopic visual sensor. A comprehensive modular software architecture was designed. Importantly, robot-dependent packages have the same standard interface, so that test beds can be interchanged with no software alteration in order to ease reproducible research. This way, cognitive abilities and hypothesis-driven feedback of the Two!Ears framework could be assessed in realistic scenarios. During the Two!Ears Summer School on Active Machine Hearing (September 2015), five MEMS-based binaural sensors were designed, and part of the architecture was ported on five low-cost TurtleBot mobile robots with system-on-chip components for binaural audio streaming (Section 4).

Guidelines to the installation of the binaural simulator and of the robotics software architecture are available in the online Two!Ears documentation (`http://docs.twoears.eu/`).

3.2 The AFE: Unified computational model of classical auditory features

The goal of the Two!Ears project was to develop an intelligent, active computational model of auditory perception and experience in a multi-modal context. The AFE represents the first stage of the system architecture and concerns bottom-up auditory signal processing, which transforms binaural signals into multi-dimensional auditory representations. The output provided by the AFE consists of several transformed versions of ear signals enriched by perception-based descriptors which form the input to the higher model stages (e.g., the blackboard described in Sect. 3.3). Specific emphasis is given on the modularity of the software framework, making this AFE more than just a collection of models documented in the literature, as, for instance, the comprehensive collection and documentation in the AMT toolbox, (Søndergaard and Majdak, 2013). Bottom-up signal processing is implemented as a collection of processor modules, which are instantiated and routed by a manager object. A variety of processor modules is provided to compute auditory cues such as ratemaps, interaural time and level differences, interaural coherence, onsets and offsets. An object-oriented approach is used throughout, giving benefits of reusability, encapsulation and extensibility. This affords great flexibility, and allows modification of bottom-up processing in response to feedback from higher levels of the system during run time. Such top-down feedback could, for instance, lead to on-the-fly changes in parameter values of peripheral modules, like the filter bandwidths of the basilar-membrane filters. In addition, the object-oriented AFE framework allows direct switching between alternative peripheral filter modules, while keeping all other components unchanged, allowing for a systematic comparison of alternative processors. Finally, the AFE framework supports online processing of the two-channel ear signals.

The AFE source code, test and demo scripts are all available from the public repository at https://github.com/TWOEARS/auditory-front-end.

3.2.1 Framework functionality

The purpose of the Two!Ears AFE is to extract a *subset* of common auditory representations from a binaural recording or from a *stream* of binaural audio data. These representations are to be used later by higher modeling or decision stages. This short description of the role of the AFE highlights its three fundamental properties:

- The framework operates on a request-based mechanism and extracts the *subset* of all available representations which has been requested by the user. Most of the available representations are computed from other representations, that is, they *depend* on other representations. Because different representations can have a common dependency, the available representations are organized following a "dependency tree". The framework is built such as to respect this structure and limit redundancy. For example, if a user requests A and B, both depending on a representation C, the software will not compute C twice but will instead reuse it. As will be presented later, to achieve this, the processing is shared among processors. Each processor is responsible for one individual step in the extraction of a given representation. The framework then instantiates only the necessary processors at a given time.

- It can operate on a *stream* of input data. In other words, the framework can operate on consecutive chunks of input signal, each of arbitrary length, while returning the same output(s) as if the whole signal (i.e., the concatenated chunks) was used as input.

- The user request can be modified at *run time*, that is, during the execution of the framework. New representations can be requested, or the parameters of existing representations can be changed in-between two blocks of input signals. This mechanism is particularly designed to allow higher stages of the whole Two!Ears framework to provide feedback, requesting adjustments to the computation of auditory representations. In connection to the first point above, when the user requests such a change, the framework will identify where in the dependency tree the requested change starts affecting the processing and will only compute the steps affected.

3.2.2 Architecture

Many different auditory models are available that can transform an input signal into an auditory representation. The actual design challenges behind the AFE framework arose from the multiplicity of supported representations, the requirement to process continuous signals in a chunk-based manner, and the ability to change what is being computed at run-time, which allows the incorporation of feedback from higher processing stages. In addition to these three constraints, the framework will be subject to frequent updates in the future (e.g., adding new processors), so the expandability and maintainability of its implementation should be optimal. For these reasons, the framework is implemented using a modular object-oriented approach.

As emphasized previously, the framework is request-based: the user places one or more requests, and then informs the framework that it should perform the processing. Each request corresponds to a given auditory representation, which is associated with a short name tag. In general, the AFE feature extraction requires the following two objects:

- A data object, in which the input signal and the requested feature representations are stored.
- A manager object which takes care of creating the necessary processors as well as managing the processing.

The manager

Most auditory representations will depend on another representation, itself being derived from yet another one. Thus, there is a chain of dependencies between different representations, and multiple processors will be required to compute a particular output. Given a user request, the manager is responsible for automatically instantiating the correct processors and signal objects, ordering the processing and routing inputs and outputs between processing stages.

3.2.3 Processors

Processors are at the core of the AFE. Each processor is responsible for an individual step in the processing, i.e., going from representation A to representation B. They are adapted from existing models documented in the literature such as to allow for block-based (online) processing.

Each individual processor that is supported by the AFE can be controlled by a set of parameters. Each parameter can be accessed by a unique name tag and has a default value. A parameter helper functionality has been implemented to display a list of controllable parameters for each individual processor. A list of all supported processors is shown in Tab. 3.1.

Table 3.1: List of supported processors.

Name tag	Description
time	Time domain signal
filterbank	Gammatone filterbank output
filterbank	Dual-resonance non-linear (DRNL) filterbank output
moc	Medial olivo-cochlear (MOC) feedback
innerhaircell	Inner hair-cell (IHC) envelope
adaptation	Adaptation loop output
autocorrelation	Auto-correlation function (ACF)
crosscorrelation	Cross-correlation function (CCF)
amsFeatures	Amplitude modulation spectrogram (AMS) features
gabor	Gabor features extraction
itd	Interaural time difference (ITD)
ild	Interaural level difference (ILD)
ic	Interaural coherence (IC)
offsetMap	Offset map
offsetStrength	Offset strength
onsetMap	Onset map
onsetStrength	Onset strength
pitch	Pitch estimation
precedence	Precedence effect
ratemap	Ratemap extraction
spectralFeatures	Set of spectral features

A demonstration of three important and frequently-used processors is shown in Fig. 3.4. A speech signal shown in the top left panel is passed through a bank of 16 gammatone filters spaced between 80 and 8000 Hz. The output of each individual filter is shown in the top right panel. The bottom left panel shows the influence of the inner hair-cell (IHC) processor. Whereas individual peaks are resolved in the lowest channels of the IHC output, only the envelope is retained at higher frequencies. The bottom right panels shows the output of

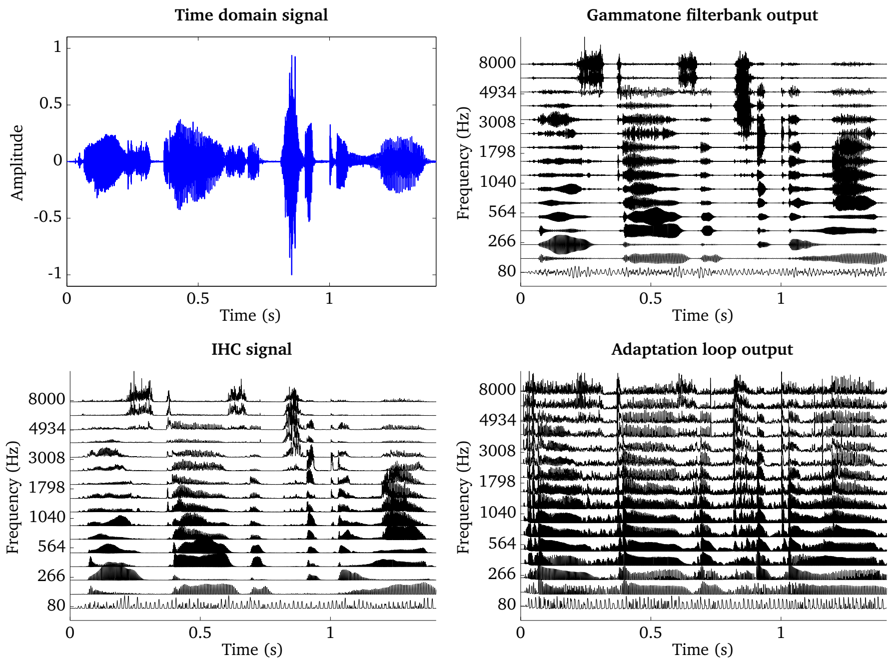

Figure 3.4: Visualization of the output of three processors in response to a speech signal shown in the top left panel: gammatone processor (top right), IHC processor (bottom left) and adaptation loop processor (bottom right).

the adaptation loop processor, which simulates the adaptive response of auditory-nerve fibers. As a consequence, abrupt changes in the input result in emphasized overshoots followed by gradual decay to compressed steady-state level.

3.2.4 Top-down feedback

A key concept of the AFE is its ability to respond to feedback from the user or from external, higher stage models. Conceptually, feedback at the stage of auditory feature extraction is realized by allowing changes in parameters and/or changes in which features are extracted at run time, i.e., in-between two chunks of input signals. In practice, three types of feedback are supported:

- a new request is placed (e. g. request an additional feature representation)
- one or more parameters of an existing request are changed
- a processor instance has become obsolete and is deleted

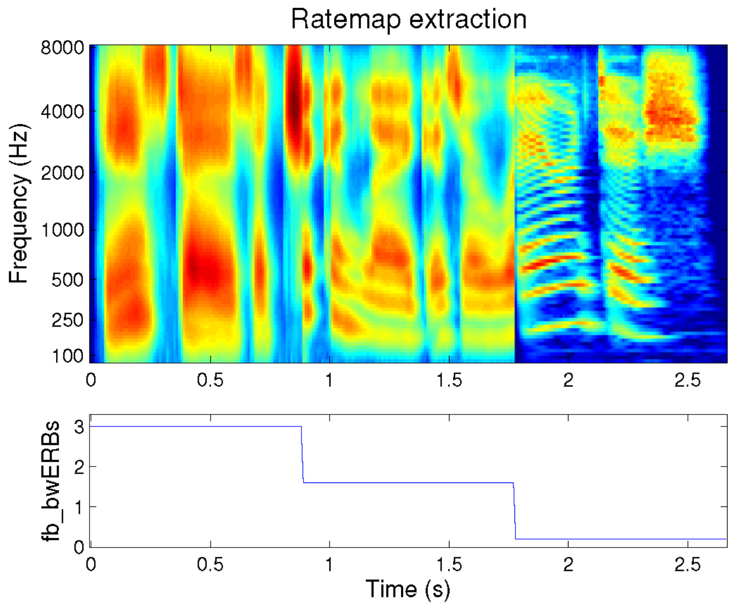

Figure 3.5: Sharpening the frequency selectivity of the ear by means of feedback

The feedback capability of the AFE is illustrated in Fig. 3.5, in which a ratemap representation of a speech signal is shown. The bandwidth of auditory filters in the original request was set to 3 equivalent rectangular bandwidth (ERB), an abnormally large value in comparison to a normal-hearing frequency selectivity. Throughout the processing, the bandwidth of the filterbank is reduced to 1.5 ERB at around 0.9 s and even further to about 0.25 ERB at 1.75 s. Figure 3.5 illustrates how the narrower auditory filters reveal the harmonic structure of speech.

In a similar way, most parameters of the other processors listed in Tab. 3.1 can be adjusted during run-time, making the AFE a very flexible and dynamic framework for extracting auditory features.

3.2.5 Conclusion

The AFE represents the first stage of the Two!Ears system architecture and performs bottom-up auditory signal processing by transforming binaural signals into multi-dimensional auditory representations. Bottom-up signal processing is implemented as a collection of processor modules, which are instantiated and routed by a manager object. A variety of processor modules is provided to compute auditory cues such as ratemaps, interaural time and level differences, interaural coherence, onsets and offsets. An object-oriented approach is used throughout, giving benefits of reusability, encapsulation and extensibility.

In fact, a considerable part of the work has been focused on the architecture and the software implementation, because basically, no new peripheral model stages had to be developed, but rather, a great variety of existing modules had to be redesigned to fit into the object-oriented structure. The gain of this software-engineering effort is only for a certain part visible in the evaluations which are done within the consortium. Through the public visibility and availability of the project and its software resources, we foresee a major step forward also in the future, when new projects initiated by the consortium members, but more importantly, the worldwide community of auditory, room acoustic and audio signal processing experts are using these modules.

3.3 Machine learning approaches for binaural auditory and cognitive modelling

At the heart of the Two!Ears project is a software architecture based on a "blackboard system", which fuses prior knowledge with the currently available sensor input. Importantly, the multi-layered structure of the blackboard system allows for modelling of cognitive behaviour in which top-down predictions (e.g. about the kind of sound sources that are present, or the kind of environment that the system is in) can influence lower levels of processing. Within this framework, we have applied state-of-the-art machine learning methods – particularly deep learning – to computational hearing tasks. This has allowed us to leverage highly effective techniques from the field of automatic speech recognition, and also allows for interesting new approaches such as combined localisation and identification of several sound sources.

3.3.1 Blackboard architecture

We developed a software architecture that integrates prior knowledge and active behaviour from a set of functional modules. These modules can work on different levels of abstraction, independently from each other or interacting, in a bottom-up or top-down manner. A key feature of our architecture is its ability to evolve, enabling easy modification, exchange and extension of modules.

The **blackboard system** integrates the framework core parts and functional modules called *knowledge sources*, and is responsible for constructing and setting up the system. It interfaces with the robot or binaural simulator – whichever is used to provide the ear-signals and command movement. It also integrates the AFE described in Sect. 3.2, which is responsible for transforming the ear-signals into auditory features needed by knowledge sources.

The knowledge sources define which data they need for execution and which data they produce. The blackboard system provides the tools for requesting and storing this data, but does not care about the actual contents. Similarly, knowledge sources are decoupled and have no "knowledge" of each other. Instead they make requests to be triggered upon firing of particular events. Events in turn can be fired by knowledge sources. The blackboard system is ignorant *a priori* of which events exist. It starts to monitor events upon receiving requests made by knowledge sources, done through the **blackboard monitor**.

The **blackboard** itself forms the central data repository of the platform. It stores knowledge sources and any shared data, in particular output of the knowledge sources (e.g., estimates of the location of a sound source). It is accessible for all knowledge sources; and it not only stores current data, but keeps track of the history of this data in order to enable knowledge sources to work on time series data. The blackboard is flexible with regard to data categories, which do not have to be hard-coded into the system.

The **scheduler** is the component of the blackboard system that actually executes the knowledge sources – after deciding the order in which knowledges sources waiting in the **agenda** are to be executed. This order is rescheduled whenever conditions determining the order may have changed, or when new knowledge sources may be present in the agenda that are more urgent. Knowledge sources with higher priority are executed before knowledge sources with lower priority. Priority can be put on knowledge sources dynamically, and can be propagated down the dependency chain of a knowledge source.

The Two!Ears system is an *active* system that works both in a signal processing bottom-up manner, and also in a "cognitive" top-down manner. Modules can change the system setup at runtime through dynamic module instantiation, registration and removal combined with on-the-fly rewiring of the communication links between modules.

Our platform is open to the public for application and further development, as a component of the open-source Two!Ears system. The system is designed to be usable and easy to configure for different tasks, code units are kept small and understandable, and object-oriented principles are used. The software can be downloaded from the project web site at `http://www.twoears.eu`. In the following, some scientific key achievements reached with the system will be outlined.

3.3.2 DNN-based localisation with head movements

Within the Two!Ears system, we developed a state-of-the-art approach that exploits deep neural networks (DNNs) for robust localisation of multiple sound sources (Ma *et al.*, 2015c).

We employed DNNs to map binaural features, obtained from the auditory front-end (AFE), to the corresponding source azimuth. Two binaural features, interaural time differences (ITDs) and interaural level differences (ILDs), are typically used in binaural localisation systems (Blauert, 1997). However, instead of estimating the ITD, the entire CCF was used as input features to the DNN.

DNNs were used to map the binaural features to corresponding azimuth angles so that posterior probabilities of there being a source at each azimuth can be estimated. A separate DNN was trained for each frequency band. This is illustrated in Fig. 3.6, where the left panel shows the CCF used as input features to the localisation DNN, and the right panel shows the posterior probabilities the DNN produces for each frequency band. Employing frequency-dependent DNNs was found to be effective for localising simultaneous sound sources, even though the networks were trained on single-source data.

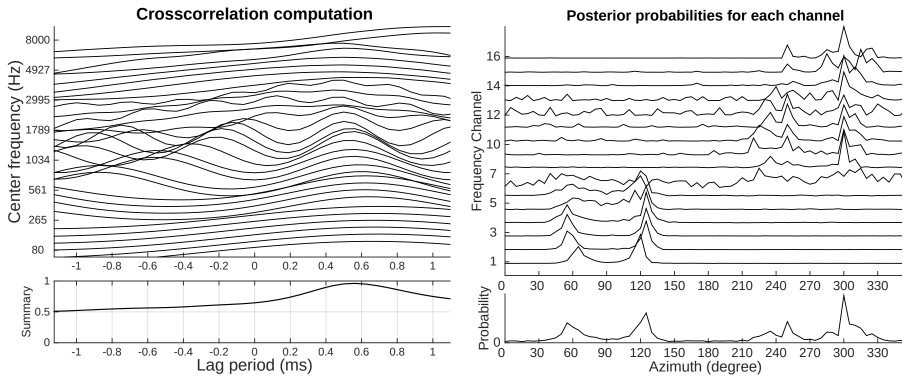

Figure 3.6: Illustration of the cross-correlation function (CCF) as input feature to the localisation deep neural network (DNN) and posterior probabilities the DNN produces for each frequency channel. Here the stimulus is a mixture of a telephone ring located at 300° and a speech source located at 120°. The telephone ring dominates the high frequency regions while the speech source dominates the low frequency regions. The integrated posterior distribution shown at the bottom of the right panel clearly exhibits large peaks at the two locations.

The DNN-based localisation framework is also able to benefit from multi-conditional training (MCT) and the use of head rotation in resolving front-back confusions in a full 360° localisation task (Ma *et al.*, 2015c). Various strategies for exploiting head movements have been explored, including an attentional system — the Head Turning Modulation (HTM) model — that turns the head towards salient changes in the environment that are relevant to the task, while minimising other unnecessary head movements.

Our experiments show that the DNNs were able to exploit the rich information provided by the entire CCF, and thus substantially reduced localisation errors when compared to a Gaussian mixture model (GMM)-based system. The MCT method was effective in combatting reverberation, and allowed anechoic signals to be used for training a robust localisation model that generalised well in unseen reverberant conditions. It was also found that ILDs were important for reducing front-back confusion errors when localising sources in reverberant rooms. The use of head rotation further increased the robustness of the localisation system, with an average localisation accuracy of 96% when multiple talkers and room reverberation were present.

3.3.3 Top-down localisation using source models

Machine hearing systems for answering 'what' and 'where' questions are typically much less tightly-integrated than they appear to be in biological hearing. We addressed this issue in Two!Ears by developing a machine hearing system for binaural localisation that exploits top-down knowledge about the source spectral characteristics.

We trained a GMM for each sound source. The source spectral characteristics were modelled using log-compressed ratemap features computed by the AFE. A ratemap is a spectro-temporal representation of auditory nerve firing rates, extracted from the inner hair cell output of each frequency channel by leaky integration and downsampling. The parameters of GMMs were estimated using the expectation-maximization (EM) algorithm separately for each sound source. Under the *log-max* approximation (Varga and Moore, 1990), the observed log-compressed energy of all present sources combined is approximately equal to the maximum energy of individual underlying sources in each time-frequency (T-F) bin. This allows us to estimate the probability of each T-F bin being dominated by a sound source in the log-compressed ratemap feature space.

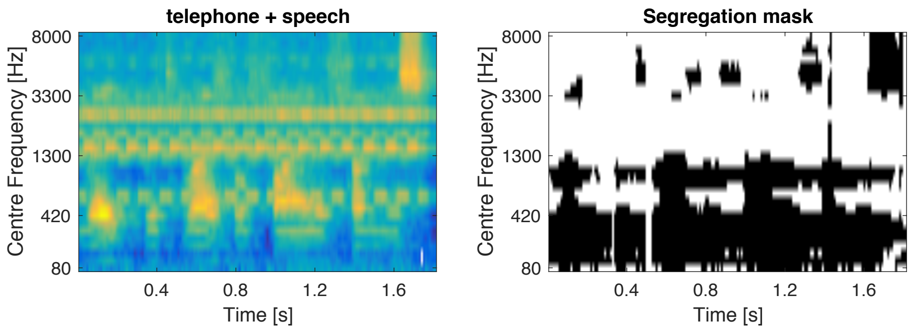

Figure 3.7: Illustration of a segregation mask estimated using top-down source models. The left panel shows the ratemap representation of a signal recorded in a real room when both telephone ring and speech sounds are present. The estimated segregation mask in the right panel shows the regions dominated by the speech source (black).

The output from this computational framework can be considered as a probabilistic segregation mask for a target source, as illustrated in Fig. 3.7. Here, the signal was recorded in a real room, using the aforementioned example of a telephone ring and a speech source, as well as some room reverberation. The left panel shows the ratemap representation and the right panel shows the regions dominated by the speech source (black). The segregation mask was estimated by using the speech source model as the target against a universal background model.

The segregation mask can be used as a weighting factor for selectively weighting the contribution of binaural cues from each T-F bin in order to better localise the attended source in the presence of background sources. As a result, top-down and bottom-up information interact within a single computational framework. Evaluation using six interfering sources with varying spectro-temporal complexity showed that by jointly exploiting top-down source models in the acoustic scene, sound localisation performance can be improved substantially under conditions where multiple sources and room reverberation are present (Ma *et al.*, 2015a).

3.3.4 Joint localisation and identification

An ultimate goal of the Two!Ears project is to form auditory objects, which requires the fusion of attributes such as source type and location that belong to a same source. In order to bind attributes to form coherent objects, we developed two approaches: (1) an extension of our identification system to operate on streams *segregated* from the ear-signals by masking the T-F space according to estimated source locations, and (2) a new system that *jointly* estimates the type and location of objects from the ear-signals.

In the first approach, sound source azimuths were first estimated using the DNN-based localisation system described above. The posterior probabilities for a set of azimuth angles were clustered based on a mixture of von Mises distributions (Banerjee *et al.*, 2005). A probabilistic masking step based on Gaussian observation models is then applied for each frequency channel to estimate a probabilistic mask for each source identified. These masks were applied to all auditory representations that can be masked in the time-frequency domain. Finally, the masked auditory features were used to retrain source identification models which are used to identify the type of a localised sound.

In the second approach, a convolutional neural network (CNN) is defined to learn a representation for the joint identification and localisation of auditory objects using environmental sounds. For each azimuth of the discretised azimuth space, one of the source types considered is placed once in each of the 72 azimuth bins. The convolutional architecture provides implicit regularisation by reducing the number of weights that need to be optimised as well as spatial invariance (Lecun *et al.*, 1998). Rectified linear units (ReLU) are used as the non-linear transfer function between successive layers. Max pooling is used for selecting the highest response within a local patch to propagate further to the next layer in the network. The output or decision layers of the network produces a two-dimensional response, in which each output neuron indicates whether a particular class is located at a specific azimuth bin.

The output from either approach described above is considered independently for each block of signal such that the temporal and spatial context is discarded. This can be improved by integrating object information over both time and space. First, we imposed a spatial Gaussian probability distribution onto the output hypothesis for each block. Summing the resulting probabilities over hypotheses creates a distribution over the whole azimuth range for each sound class. These probability maps are finally integrated over time to give robust estimates of the location and source identity.

3.3.5 Conclusion

The Two!Ears project has made substantial progress against its ambitious goals. In complex multi-source environments, the system is able to determine the type of sound sources that are present, localise them and track them. The localisation and attentional systems are able to direct the head of the robot in order to resolve ambiguities. In some cases, such as when determining the azimuth of a target voice in a mixture of four other voices, as in the experiment of Kopco *et al.* (2010), we have shown that the performance of the system approaches human performance (Ma and Brown, 2016).

In many of these cases, top-down feedback in the system has been shown to be vital. For example, the performance of source localisation can be significantly improved if top-down feedback is used about the spectral characteristics of the sound sources that are present. Similarly, top-down information gleaned about the number of sources in the environment, which can be done from acoustic models, can provide important cues to lower-level processing such as stream segmentation.

A notable aspect of the system is the way in which it leverages statistical machine learning within a blackboard architecture. Various aspects of the Two!Ears system make use of graphical models and deep neural networks, which provide powerful tools for exploiting the acoustic data available. Multi-conditional training across a wide range of acoustic conditions has been shown to improve the performance of the system in a number of respects, including source localisation and classification of sound type.

3.4 Auditory and crossmodal feedback

3.4.1 Reflexive head turning (turn-to-reflex) and sensomotoric feedback to enhance speed and accuracy of sound-source localization

As regards reflexive feedback, we have worked on the so-called **Turn-to-Reflex**. This is the reflex of turning the head into the direction of the sound source in the case that a sudden sound pops up. The reflex helps putting the sound source into the visual field and thus identifying the reason for the sound. Interestingly, blind people do usually not turn the head in a frontal position to the source, but rather slightly to the side, namely, in the direction of best hearing[3]. Whether this is still pure reflexive is worthy of discussion. In any case, appropriate head turning requires fast and reliable auditory source localization, also in situations with concurrent sources. In Two!Ears we have mastered this task. Actually, different reflexive approaches have been implemented and tested to this end, such as

- Gaussian-mixture-based (GMM) localization
- Deep-neural-network-based (DNN) localization (Sect. 3.3.2, this document)
- Localization by "cophase-and-substract" the left and right ear-signals (null-antenna approach, see D4.3, Sect. 4.2.4)

If, in the course of sound localization, the commands that cause head movement are considered in addition to the auditory cues, localization becomes significantly faster and more precise. It seems that this happens subcortically in humans in a reflexive process. In Two!Ears we have modeled this **sensorimotor-cue processing** as a three stage process (D4.2, Sect. 2.8; D4.3, Sect. 4.2.7)– Fig. 3.8.

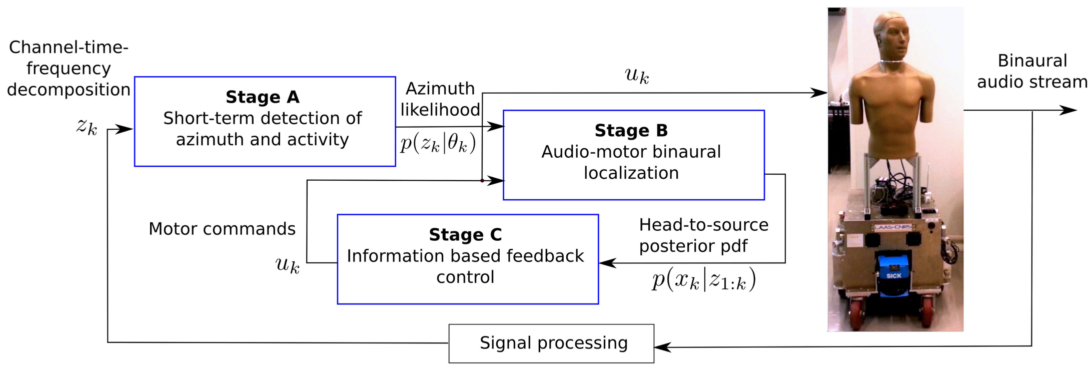

Figure 3.8: Three-stage active sensorimotor binaural sound source localization

3 observed by the authors

3.4.2 Signal-specific control of the ear-filters ("sharpening the ears") for improved sound-source identification and localization

In an interesting Two!Ears experiment it was shown that the Two!Ears system can be made to "sharpen its ears" by **signal-adapted control of its earfilters**. To this end, the system tries to recognize whether an incoming sound signal belongs to a class of signals that the system has learned beforehand. If this is the case, adequate enhancement filters are applied to the signals, to the end of enhancing binaural cues and, consequently, sound-source-class identification and localization. As the system uses prior knowledge regarding the signal classes to consider – that is, source models – this is a reflective feedback process (Sect. 3.3.3, this document; D4.3 Sect. 4.2.1).

Also the filter features of the external ears, i.e., their HRTFs (head-related transfer functions) can be exploited to improve localization and to allow for segregation of competing sound sources – another kind of "sharpening the ears". Therefore the two ear signals originating from a certain *desired* sound source are de-convolved in respect of the adequate HRTFs, and then co-phased and subtracted. What is left are all signals that do not stem from the desired source and are thus taken as noise. Subtracting the noise from the total signal mix lets the desired signal stand out clearly. This approach can also be used to build **Cocktail-party processors**, namely, systems that segregate competing sound sources. Since the directions of sound incidence and the HRTFs have to be known by the system, the processes are knowledge based an thus reflective (D4.3 Sect. 4.2.3). The algorithm works best when the number of sound sources is known and the head (resp. dummy head) is allowed do move freely.

3.4.3 Head-position and -motion control based on temporal (attention) and spatial (precedence) salience

Temporal saliency Head turning is not necessarily a purely reflexive process but can also be controlled reflectively, based on the understanding of the scene concerned or at least of relevant parts of it. This raises the question of salience and attention. As an example, an algorithm for exploiting temporal saliency in the context of head movements has been developed in Two!Ears. The basic idea is the following. If a sound source pops up in an environment for the first time, it is considered relevant and worthy of turning the head to. However, if a same or a similar (i.e., congruent) one appears, as compared to the prior one, it is considered less interesting because it is known already and, thus, head turning does not provide any advantage. Yet, if a new sound source enters the field that significantly differs from the prior one(s) (i.e., is incongruent), head turning is initiated again. The model algorithms have been amended to include visual events in the object-building process, that is, allowing for multimodal fusion if useful (D4.3

Sects. 4.3.1;4.3.2;4.3.6).

Spatial saliency As already mentioned above, **localization accuracy** in single- and multiple source environments can be significantly increased by reflective head movements. For example, it has been shown for two sources at different angles with respect to the listener that the classification of sound sources can be significantly improved. The best performing spatial arrangements have been found with a feature-based machine-learning approach (D4.3 Sect. 4.2).

A psychoacoustic phenomenon that does not only comprise reflexive but also reflective elements, is the so-called **Precedence Effect**. In general terms, it describes that in situations where a sound wave that comes directly from a source but is accompanied by delayed copies of it, such as wall reflections, the auditory event is formed in the direction of the first incoming wave front – the "direct sound" – Fig. 3.9 In Two!Ears,

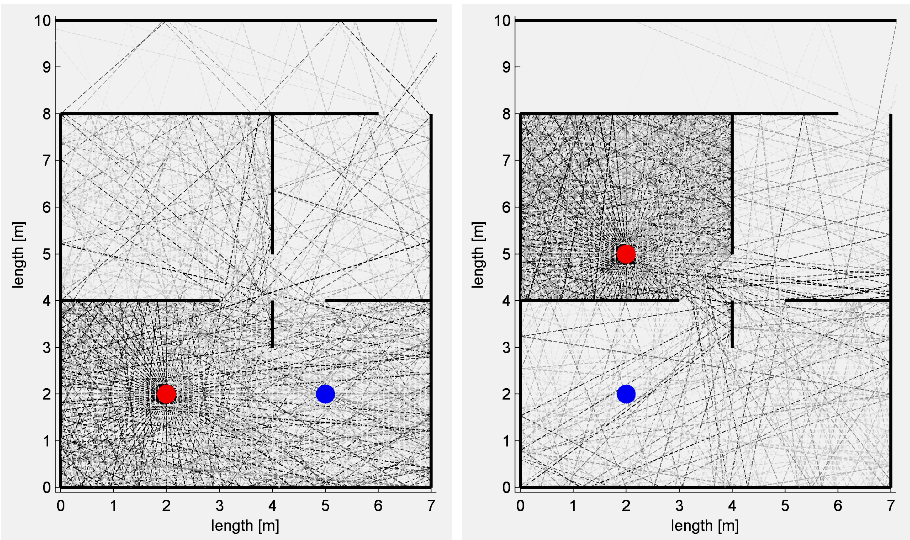

Figure 3.9: Ray-tracing simulation of the sound field in the Adream lab (i.e., the computer apartment set up at LAAS in Toulouse) as a means for identifying the direct sound and the reflected sounds. Left: Scenario with a non-occluded sound source. Right: Scenario with an occluded sound source. Sound sources are depicted as red dot, binaural receivers as blue dot

a novel, correlation-based Precedence-effect model has been developed, which cannot only deal with impulsive but also with ongoing sounds. The basic idea of the model can also be applied to HRTF-based signal enhancement, de-noising and de-reverberation. Further, it could be demonstrated in Two!Ears, that the latter two applications can be enhanced by informing the assessor visually about the geometry of the room in which he/she is sitting. In other words, the Precedence effect can be influenced by cross-modal cues and the situational knowledge derived from them – in our case visual cues. The Two!Ears model takes account of these facts (D4.3 Sect. 4.2.2). Related algorithms have been developed in Two!Ears that are able do perform De-reverberation (D4.3 Sect. 2.4).

3.4.4 Cognitive scene analysis (incl. meaning assignment) and action control in search-&-rescue scenarios (D4.3 Chap. 5)

Dynamic auditory scene analysis (DASA) requires decisions on a cognitive level, for example, when assigning meaning to scene elements and/or interpreting scenes to induce appropriate actions. To this end, feedback from the cognitive level to the auditory-signal processing level had to be considered.

An efficient software package, the *Bochum Experimental Feedback Testbed* (BEFT), has been developed for this purpose and is now in use to test feedback routines of moderate complexity with cognitive elements to them. BEFT integrates the interactive binaural simulator developed in Two!Ears (see Sec. 3.1.1) for auralization and communicates seamlessly with the blackboard. BEFT enables reliable testing of basic feedback routines, with a focus on active exploration of audio-visual scenes by employing mobile sensor platforms. Visual cues are used for assistance if necessary. To achieve this, the construction of a variety of novel classes and knowledge sources (KSs) was mandatory.

Each virtual AV source is specified by a set of high-level information that describes cognitively-meaningful characteristics of the represented physical source. These *meta tags* currently include the *category*, the *role*, and the *gender* of the AV source. So far, the employed categories are {human,threat,alert}, with roles being {victim,rescuer,fire,siren}, and gender chosen from {male,female,NA}, where "NA" means that the attribute gender is "not applicable", for instance, in case of inanimate scenario entities. The meta tags allow for cognitive analysis and understanding of the observed scenario. Based on them, the virtual robot can, for example, be asked to "find all female humans that are classified as victims". Meta tags can principally be inferred by appropriate classifiers from incoming audio-visual features. However, in the current state of the system, the tags are mainly based on ground-truth information available for the scenario concerned.

Source localization in the Two!Ears system is based on auditory-feature extraction using Two!Ears's AFE (see Section 3.2). Azimuth estimation is performed by either a standard GMM-based or a deep-neural-network-driven approach. To enable these steps, ground-truth scenario information can be employed if adequate, thus avoiding additional computational costs for auralization/auditory feature extraction. Further, this strategy allows to operate with arbitrary source characteristics, even when there are no actual sound samples available representing particular source identities, categories, roles, or gender.

BEFT uses continuous multi-source triangulation techniques to render *environmental maps*. With the GMM as defined above, environmental-map formulation is straightforward. Namely, the GMM components are projected into the x/y–domain, there initiating growing peaks at the positions of the active sound sources. Based on the entries in the short-term memory as they stand, the robot can already perform some basic cognitive tasks, such as "find all male humans in the given scenario". To that end, the system will build up two *task-specific maps*. The first map augments the peaks in the GMM-based environmental map with the corresponding auditory objects' probabilities of "being human". The second task-specific map is formed in a similar way, namely, by using the "male" meta tag for augmentation. The two maps are then merged by multiplication. Novel identities, categories, or roles can readily be integrated into this scheme and will allow to address cognitive tasks of significantly higher complexity.

In summing up, the proposed *Bochum Experimental Feedback Testbed* (BEFT) allows to perform basic experiments with regard to active exploration, and multi-modal feedback in the context of the Two!Ears system. Continuous multi-source triangulation works well as long as the environment is of moderate complexity, but there is headroom for improvement. As a result, comprehensive environmental maps are set up. These maps, on the one hand, allow to visually understand the virtual robot's behavior. On the other hand, they provide the basis for cognitive analysis of the processed scenario. As the analysis methods are assessed in a virtual environment of BEFT, experiments can be conducted without having to set up real-world search-and-rescue scenarios, thus, among other benefits, preventing humans and devices from potentially dangerous situations – for details see (D4.3 Chap. 5).

Proof of Concept In search-and-rescue (SAR) scenarios a victim in a (moderately) complex environment is identified, localized and rescued. The processes and actions are predominantly based on binaural cues, derived from the ear signals of a virtual robotic platform that can actively move about in the scene to be explored. Data that are not reliably available from the auralized scenario system, are emulated in the basis of ground-truth data. In this way the full functionality of the Two!Ears system is demonstrated. Visual cues are employed for assistance if necessary. For this purpose, the virtual robot is equipped with a camera (D4.3 Chap. 5). For demonstration, a simplified version of the Adream lab in Toulouse has been replicated. A scenario is simulated that starts with a normal lab situation but suddenly evolves into a catastrophic situation. After an assumed explosion, attendant lab employees turn into either victims or rescuers. A little later, a fire starts in one corner of the lab. The robotic agent enters the scenario and actively explores the terrain in order to infer positions of all animate entities. This is primarily

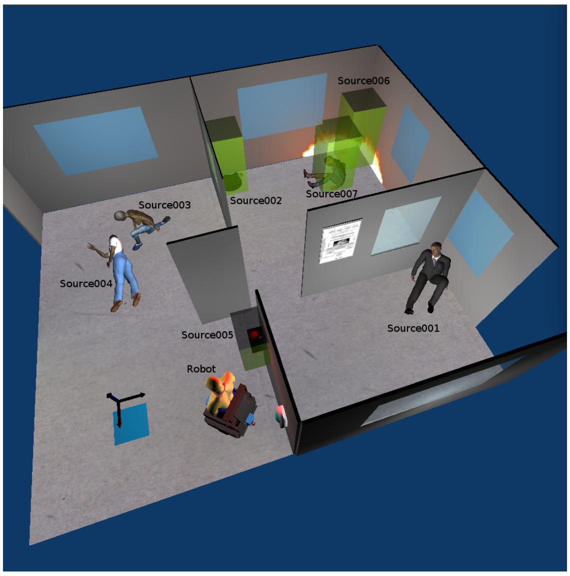

Figure 3.10: Demo experiment with BEFT. **Greenish boxes** indicate active sound sources, the **blueish rectangle** is the idle position of the robot. In the depicted situation, the robotic agent will first rescue the prioritized entity with label "Source007"

done based on auditory cues. Consequently, the robot triggers an audio-visual alert to warn potential co-workers. Thereafter, it evacuates all potential victims as recognized to be present in the environment. The evacuation follows a rescue plan that the system develops by itself, based on its knowledge of the environment and the behavior of the (virtual) persons and other entities involved. Figure 3.10 presents a bird's-eye view of the sample scenario.

4 Impact and dissemination

TWO!EARS has targeted visionary research on advanced computational models of human audition. The project has been interdisciplinary in several ways, and has tried to put into practice what has been portrayed as the way to follow in audition modelling, namely, to include cognition and the extraction of meaning, and the active way in which humans explore their environment. Given the overall strong role of sound in our modern society – including both positive aspects, like communication, sound design, sound scape generation, and negative aspects such as noise impact, stress, hearing damage – it is of great practical value to have automatic tools for the analysis and evaluation of sound-related effects. With our scientific approach and dissemination strategy we believe to have reached the ambitious goals of TWO!EARS expected to have breakthrough character and show strong impact on science, technology and society.

There are a number of results that represent exploitable foreground for the project. Since TWO!EARS aimed at reproducible research in an Open Science context, there were deliberately no patent applications produced, or other trademarks or products developed that would target a direct selling of the work. Instead, the foreground may be exploited by anyone active in any of the areas addressed by the TWO!EARS topics, be it academic institutions to build on our research or industry who wish to use our framework for their corporate R&D or product developments.

The exploitable results are (see also `www.twoears.eu` and `https://github.com/TWOEARS`):

a) Database of acoustic scenes and measurements
b) Auditory front-end (AFE) software
c) Blackboard software, knowledge sources and associated software framework for machine learning
d) Robotic unit for KEMAR head rotation and stereoscopic vision (mechanical design and software)
e) Robotics software and bridge to auditory front-end / blackboard software
f) Development guidelines for low-cost Turtlebot solution for active exploration in an educational context
g) Virtual test environment for cognitive functions (LVTE)
h) Test methods for quality evaluation of high-end audio systems
i) Software components for audio quality assessment

Impact

The research conducted in Two!Ears has a number of prominent key-features:

Human-centeredness Two!Ears has addressed audition modelling that has two types of applications, (i) those that directly concern and support human auditory perception and communication capabilities – as in hearing aids and telecommunication technology, and (ii) those that benefit from models of auditory perception to fulfil complex tasks that so far have been out of reach for technical systems – for example, auditory scene analysis for robotics, speech understanding in adverse conditions or automatic annotation of acoustic scenes for personal diaries or for an active adaptation of the acoustic environment.

Toolbox of evaluated modules Two!Ears has put audition modelling into more complex and realistic contexts than previous related activities. The resulting Two!Ears system consists of modules that are individually improved and extended according to the needs that emerge from embedding them in an integrated system.

Active Hearing Two!Ears brings together bottom-up auditory-feature extraction with cognition modelling and active exploration for object identification and meaning assignment, rigorously modelling the entire audition 'chain'. This has long been considered as one of the main paths to follow for more prominent advances in hearing and audio technology.

Cross-modal integration The employed architecture enables inclusion of vision- or video-related as well as listener- and head-position information, in this way providing a new, perception-inspired approach for cross-modal data-fusion and an improved system for audiovisual scene analysis.

Meaning extraction and awareness It was a main project goal that the Two!Ears system should be able to identify objects and extract meaning. Hence, it principally is *content-*, *environment-* and *user-aware*, based on learned internal references.

Exploration & top-down adaptation Based on the system's awareness and its ability to translate this awareness into active exploration and top-down feedback, the individual system components and the integrated system can flexibly be adapted to the application domain, the selected environment or the hearing characteristics of an individual user – and be implemented so as to automatically optimise its own robustness and performance.

Public model software & audiovisual scene database The Two!Ears model software has been made available Open Source. Further, the Two!Ears database of recorded and synthetically created, and subsequently (multi-layer) labeled acoustic and audiovisual scenes have been made publicly available as well, together with tools to access it. Thus, both the model software and the database can be used as an unprecedented resource for reproducible science as well as for further analysis and research.

It is assumed that the ambitious Two!Ears goals may have a transformational impact at different levels. With the interdisciplinary and holistic approach, Two!Ears is expected to have significant *scientific impact*. With the range of applications that are likely to benefit from or to be enabled by Two!Ears, it is clear that the project will have an *impact on technology* at a European and global scale. A direct short-term impact is expected for components of current and near-future task-specific audio QoE models, or the pre-processing components of hearing-aids. The main impact of Two!Ears on ICT at the European and global scale is expected for the mid- and long-term future: With its foundational combination of bottom-up with top-down processing including cross-modal integration and feedback, the Two!Ears system is a versatile tool for aural-scene analysis, likely to result in significant improvements of audio-transmission & -reproduction, hearing-aid-technology, robotic audition, and in improved front-ends especially for distant speech recognition systems, getting us closer to solving the cocktail-party problem using computers.

Two!Ears is human-centered: The majority of applications or services that are likely to benefit from its results refer to human auditory perception or communication and are typically designed for improving the quality of life or work of their prospective users. Thus, as a result of the impact at scientific and technological level, Two!Ears is expected to have a positive *impact on society*.

Scientific impact
The Two!Ears system and its components provide a proof of concept for a variety of uni-disciplinary as well as inter-disciplinary innovative ideas that can be built on in future research and development. The transformational impact will thus apply to the individual domains, their combinations, and along multiple dimensions.

In contrast to the previous approaches of machine-learning or automatic pattern recognition applied to acoustic signals, rigorous perception-modelling has been used to provide bottom-up input features. This is a considerable transformational step from, for instance, microphone-based signal processing information used in object-localisation tasks (Gustafsson *et al.*, 2003, Lombard *et al.*, 2009, Teng *et al.*, 2010), or the more restricted perception models used, e.g, for stereophonic-audio-reproduction evaluation (Rumsey, 2010).

In order for machine-learning technology to scale, learning methods must become autonomous, in other words, inductive learning with no or marginal user interaction must be feasible. Drawing conclusions from incomplete or contradictory data can be supported by the intended coupled graphical-model architecture, which incorporates rule-based and statistical knowledge in a system that supports reasoning not only about the unknown variables of interest, but also about its own state of knowledge – thus enabling an exploratory behaviour and (more) autonomous learning. A dynamic blackboard architecture has been realised on this basis, enabling cooperative behaviour of agents based on a common set of structured *knowledge sources* – which comprise knowledge about the structure and composition of auditory scenes and sources in this case, including vision-related information for cross-modal integration. This integrated approach is an absolute novelty in the field of auditory modelling and appears to be a promising approach to flexibly build future applications on. The successful usage of this system in conjunction with modern machine learning approaches such as DNNs has been proven by Two!Ears for example in Ma *et al.* (2015c) and (Ma and Brown, 2016).

With the inclusion of suitable feedback mechanisms for top-down refinement of the extracted meaning in terms of active listening, both a task-dependent module selection and a 'sharpening-the-ears' module have been considered. By including top-down audiovisual feedback, adaptation at different levels is feasible. In general, such a functionally adequate awareness and adaptability of an auditory perception model or 'agent' system paves the way for a large variety of future scientific activities and the development of new applications.

This active exploration- and feedback-based approach is further enhanced with its implementation on a mobile robotics platform. This way, actual human-like exploratory head-movements as well as translatory displacement are enabled, bringing the overall system even closer to how humans explore their environments. This explicit inclusion of robotics ensures the optimal use of bottom-up auditory perception models, advanced machine-learning and graphical-model architectures as well as top-down feedback mechanisms for robot audition. In return, it provides an ideal platform enabling modes of exploration so far impossible for auditory perception and cognition models, and a test-bed framework to evaluate and optimise model components as well as entire model systems. As this framework has been made publicly available, it is certain to foster new science in this field, and based on reproducible results. Moreover, the Two!Ears scene database serves as a rich resource for further research and model- and robotics-system development. The scientific impact at this level has been ensured with the strong links of the consortium to the psychoacoustics and hearing modelling, machine learning and robotics communities.

Technology impact

The Two!Ears model framework is targeted as the front-end for a great variety of applications, and is thus expected to have a high impact on ICT at large. Some of the possible applications of the Two!Ears' results will be discussed in the following.

Hearing aid technology

The Two!Ears system may have a direct impact on future *hearing-aid development*, fostering the successful European hearing-aid marketplace. The Two!Ears approach is well adapted to the two most important current trends: Using fast communication between the two devices for the two ears finally enables binaural processing to be used. This way, a significant gain in source separation can be achieved. Different companies such as *Oticon, Phonak* and *Siemens* are working on such topics. Here, TWO!EARS has made a rich set of research tools available that can directly feed into the hearing-aid industry's requirements. TWO!EARS can have a very significant impact here by providing tools for *R&D* towards a more powerful scene analysis. This way, hearing-impaired listeners may be much better assisted by hearing aids in solving the *Cocktail-party Problem*, such that important sources are identified and selectively enhanced. With the content-, user- and environment awareness and the additional ability of feedback-based adaptation of TWO!EARS, a hearing-aid can – in principal – be enabled to faithfully extract the information that is most relevant for the user in a given situation.

By changing bottom-up modules to simulate certain hearing deficits, it can be evaluated how these affect the whole system performance. With the additional ability to simulate individual users including their behaviour, the TWO!EARS' development system or the robotics deployment system can directly serve as a hearing-aid-algorithm-development workbench. Also, TWO!EARS enables subsequent research to evaluate whether degraded performance of humans in adverse condition is rather caused by deficits in the bottom-up, or in the top-down channels (e.g. Teki *et al.* (2011)). Reaching impact in this highly specialised and research-driven industry required vivid exchange with the respective key-players. Here, the good industry contacts of especially DTU, TUB & TUE have helped to establish awareness of the TWO!EARS results, for example in the context of the 'Oticon Centre of Excellence for Hearing and Speech Sciences' (CHeSS). Besides the transfer of project results, TWO!EARS' highly trained experts themselves are a very interesting resource for the industry to bring the project-related knowledge into applications; for example, one of the TWO!EARS junior researchers has started working at Oticon now.

Assisted-living technology
Also the development of systems for *assisted living* may be supported by the TWO!EARS modelling framework. With such systems, deaf or hard-of-hearing persons benefit from pre-processed auditory information provided to them by alternative modalities. For instance, a central system or a system worn by the user can be envisaged that converts acoustic information such as a ringing phone, a buzzing smoke detector, one's baby crying, or persons walking, into tactile or visual information, fully replacing the missing auditory cues[1]. Examples are small transmitters in doorbells, lights hooked onto telephone lines, door announcers that can also be used outside the home for indicating that someone is knocking or ringing, a visual indicator added to the smoke alarm, and so forth. Likewise, solutions have been proposed for the cars of deaf or hard-of hearing persons to indicate, for instance, approaching sirens. Components from the TWO!EARS framework are highly useful when developing such technology, as usually all natural or synthetic acoustic information is present in a form directed to normal-hearing persons. Here, a fully functional, mobile

1 Products can be found, for example, at `http://www.harriscomm.com/`. Manufacturers include *Ameriphone, Bellman & Symfon, Global Assistive Devices, Sonic Alert*

model of human audition will enable the user to behave like a user with functioning hearing in arbitrary environments. During the project, there have been first activities in this area by USFD, who will exploit outputs of Two!Ears through CATCH, their Centre for Assistive Technology and Connected Healthcare, see `http://www.catch.org.uk`. Here, there is also potential for acoustic monitoring relating to healthcare more generally, where different partners have or have applied for follow-on funding.

Distant speech recognition
Developing better *distant-speech recognition* systems was not a direct goal of Two!Ears. However, an audition model that delivers a high-performance auditory-scene analysis is a perfect front-end for such systems. Two!Ears provides a whole set of tools that can be used to enhance distant-speech-recognition performance, and results we obtained in this regard indicate the strong potential in this domain. The Two!Ears consortium has engaged with the speech recognition community, with different scientific contributions such as the attendance of the ASRU meeting in Arizona, USA (Ma *et al.*, 2015b).

Quality of Experience
Mastering the widely unsolved problems regarding prediction and control of *user experience and its quality* is an absolute necessity in ICT. A model with content-type-specific adaptation that is steered by the model itself in a top-down fashion and may possibly include internal references relevant for a certain user group, is expected to lead to a substantial impact at scientific and ICT level. The rich toolset of quality-modelling modules provided by Two!Ears was proven to principally enable developments along these lines (Wierstorf *et al.* (2015), Raake and Wierstorf (2016), modelling see `https://github.com/TWOEARS`). A full model and the world-knowledge of listeners is difficult to implement especially for high-quality audio systems, and we are certain that our software, database[2] and quality-related modelling are very attractive also for relevant R&D groups outside the Two!Ears consortium, to the extent that they will employ the established system and test their functional modules and algorithms as system components. Typical applications of such QoE models are the assessment of audio reproduction technology (such as 5.1, 7.1 or other surround-sound systems, multi-loudspeaker systems using *Wave Field Synthesis* (WFS) or *Higher-order Ambisonics* (HOA)), of the room acoustics of performance space, such as concert halls, and of spatial-audio-coding methods such as MPEG-H.

With the ability of active exploration, the Two!Ears system provides components for audio-quality assessment at different spatial positions in a given acoustic scene, and human-like active exploration, e.g. to assess the 'sweet-spot(s)' of an audio system or room. Especially new types of audio technology may suffer from new types of artifacts, which can currently not be measured with any instrumental method. Assessment with the human hearing system in listening tests is the only way to validate them. The evaluation

2 `http://docs.twoears.eu/en/latest/database/`

of QoE with binaural auditory models would not only help in the evaluation of quality, but also provide very useful information for the further development of such technology. To make them more widely known, Two!Ears results on quality assessment will be disseminated into standardization by TUIL at respective upcoming group meetings (e.g. ITU-R Working Party 6C, who tried to extend ITU-R BS.1387-1 towards multichannel audio in the past (Liebetrau *et al.*, 2010, Rumsey, 2010)). First discussions of Two!Ears' results in this area were held with representatives of companies and organizations such as Deutsche Telekom, Bang & Olufsen and Fraunhofer.

Robotics
An application area for Two!Ears is in *robotics*. With a Two!Ears-like system, robots can coherently parse the environment and orientate themselves or can respond to speech addressed to them. Since only two ears are required instead of a microphone array as it is most frequently used, such a system can be of special interest for humanoid robotic systems. With Two!Ears, we now achieve very good performance using two ears (and importantly, top-down feedback) for tasks which previously required microphone arrays to work well, for example in ASIMO[3].

The Two!Ears platform constitutes a significant complement to the two current most prominent open-source software libraries for robot audition, namely: $\mathcal{HARK}$[4](Nakadai *et al.*, 2010); and *ManyEars*[5](Grondin *et al.*, 2013). As aforementioned, its scope is broader in that it includes human-like audio streaming from any ALSA-compliant interface, low-level binaural signal-driven functions, high-level cognition and top-down hypothesis-driven feedback. Nevertheless, as it is based on the ubiquitous ROS middleware and as it is the result of a clean model-driven design, it can easily be interfaced with existing libraries. Thus, it can be expected that Two!Ears has provided knowledge with potential impact on auditory-scene-analysis systems for robotic applications.

Teleconferencing and tele-meetings
Teleconferencing and tele-meetings are gaining in importance – not least to reduce travels and foster an environmentally-friendly communication and collaboration behaviour. Such systems can only be successful when enabling high-quality interactions not missing too many essential features of face-to-face meetings. Here, spatial audio was shown to significantly improve intelligibility, listening effort and even the memorisation of who said what in a given call (Skowronek and Raake, 2015). Incorporating an exploratory hearing model based on Two!Ears can open the floor for new or improved interactive multi-media applications and related research activities – for example, enable users to efficiently participate in virtual communication scenarios or to participate in spite of hearing limitations. Intelligibility

3 Robot initially introduced by *Honda* (`http://asimo.honda.com/`)
4 $\mathcal{H}$onda Research Institute Japan $\mathcal{A}$udition for $\mathcal{R}$obots with $\mathcal{K}$yoto University, `http://www.hark.jp`.
5 `https://introlab.3it.usherbrooke.ca/mediawiki-introlab/index.php/ManyEars`.

and ease of communication are the key performance indicators for speech-communication technology. Predicting and controlling them is indispensable not only for audio and multi-media technologies at large but also, for example, in room acoustics for both real and virtual spaces. In the area of intelligibility modelling, too, Two!Ears has achieved highly promising results (Relano-Iborra *et al.*, 2016).

Research Data Management and Open Science

Both the irreproducibility of scientific results and the insufficient availability of underlying data led to a call for professional management of research data. General recommendations and best practices on professional data management within the scope of the Horizon 2020 program (European Research Commission) can be summarised as:

- develop a detailed data management plan,
- use of workflow tracking during the entire research process,
- data must be made findable, accessible, interoperable and reusable (FAIR),
- use open licensing models when publishing data and implementations, and
- offer training and qualification on data management.

In Two!Ears these principles have been implemented carefully together with a clear commitment to open science. The following tools have been used to ensure data provenance and reproducibility in all research processes

- Redmine (red) which provides a variety of valuable tools for collaborative engineering and research. Among others this includes the management of access rights, version tracked Wikis, issue management, work scheduling and a seamless integration of various version tracking systems. Redmine is Open Source Software.
- Apache Subversion (SVN) (INF, a) is a version management system initially developed for the tracking of source code. It can handle large amounts of data reasonably well since it allows for a partial checkout of data. SVN is therefore used for the version tracked storage of (binary) research data.
- git (INF, b) can be seen as an advancement of SVN. It provides versatile functionality for a highly distributed software development process. For instance advanced branching and merging features. It is however not so well suited for the handling of large amounts of (binary) data. git is therefore used for the software implementations developed and maintained.

The systematic management of primary and raw data is a prerequisite for the publication and long-term preservation of research data. Publication and long-term preservation are realized by building up a long-term citable electronic archive. Its entries are uniquely referenced by digital object identifiers (DOIs). The archive provides the grounds for extended publications with references to implementations and data, and for the use of Two!Ears results by third parties. The data is accompanied by clear licenses, e.g. from the Creative Commons (CC) framework (INT, b), in order to clarify the conditions of use.

Contribution at the European level

The ambitious goals of Two!Ears have been reached with the joint forces of researchers from different fields such as audio-signal processing and engineering, psychoacoustics, audition modelling, vision modelling, machine learning, speech recognition, knowledge & cognition modelling, neurophysiology, QoE assessment, and robotics. With our balanced mix of established scientists and young researchers, the composition of our consortium itself has guaranteed that the targeted inter-disciplinary research approach could be established with a long-term perspective in the European research landscape. The formation of a European consortium to work on this project has been a natural choice, since acoustics and auditory perception are very strong European research domains. Similarly, the hearing-aid, audio technology, and telecommunications industries are strongly represented by the respective European players. In order to recognise and benefit from the expertise of international research teams working on Two!Ears topics, we collaborated directly with RPI as a project partner, and an Advisory Board that consisted of three distinguished researchers from the auditory sciences, namely Michael Akeroyd (UK), Bill Yost (USA), and Israel Nelken (Israel).

About four years before the project started, five of the Two!Ears partners together with four further international research groups formed an intellectual circle, *AABBA* (Blauert *et al.*, 2009), to get a better understanding of the problems of modelling audition for the benefit of modern technologies. Since then, further partners have joined *AABBA*, with now eight of the Two!Ears partners being part of it. In fact, as far as modelling of binaural hearing and sound-quality assessment is concerned, *AABBA* includes the leading experts worldwide. Two!Ears has closely cooperated with *AABBA* to achieve substantial synergy and faster dissemination of results. Also after termination of the project, the Two!Ears consortium will continue the dissemination of results via collaboration within *AABBA*.

Two!Ears has been integrating well-known and novel approaches into an innovative, flexible system that models human audition. Due to its modular design, the system will be able to grow further in its evolutionary usage also after project lifetime. Aside from the model-framework, Two!Ears provides its labeled scene database as open resource. With the broad range of possible applications of the Two!Ears results, it is expected that the project will pave the way for future European and worldwide research in this domain. The continuous usage and development of the software and database beyond project lifetime requires dedicated resources by the Two!Ears consortium. The sheer maintenance of the software and database will be ensured from within the consortium. For future extensions, respective funding opportunities are currently being explored by different Two!Ears partners.

Dissemination

There have been a number of dissemination activities throughout the project. The Two!Ears website (http://www.twoears.eu) provides all relevant project-related information. The Two!Ears publications and a number of presentations are made available on the website. Further, via http://github.com/TWOEARS, the code developed in Two!Ears has been made available using current free software distribution means, serving as the dissemination platform for all further Two!Ears software development. The Two!Ears data base is available at https://dev.qu.tu-berlin.de/projects/twoears-getdata/repository. The project can be reached via Linkedin (https://www.linkedin.com/groups/TWO-EARS-6666486) and is active on Twitter (@TwoEars_eu). Project participants have organized Two!Ears-related special sessions at conferences (Forum Acusticum 2014, Krakov http://www.fa2014.pl/; DAGA 2014, Oldenburg http://2014.daga-tagung.de/de/programm/vortragsprogramm; INTERSPEECH 2015, Dresden, Germany http://www.interspeech2015.org/; International Congress on Acoustics – ICA 2016, Buenos Aires, Argentina ica2016.org.ar/, and given invited talks about Two!Ears at different events (e.g., EAA Joint Symposium on Auralization and Ambisonics 2014; Tonmeistertagung tmt 2014; DAGA 2015, Nürnberg, Germany, www.daga2015.de/; SpiN workshop 2015, Copenhagen, Denmark (http://www.spin2015.dk/), FET-Open EARS project meeting, November 2015; German Acoustics Conference DAGA, Aachen, Germany, March 2016; 140th AES Convention, see http://www.aes.org/events/140/workshops/?ID=4899; Workshop “Sensing: From minds to machines” at Ben Gurion University, Be’er Sheva, Israel; Hearing4All International Symposium, Oldenburg, 2016; invited plenary talk at DAGA 2017, Kiel, Germany; several further invitations for Acoustics ’17, Boston, USA (Blauert, Braasch, Raake, Spors, Wierstorf)).

A Two!Ears summer school was held in Toulouse, France in September 2015 on Active Machine Hearing[6]. To provide the 19 participants (plus 3 invited speakers and 13 members from the consortium) with test beds necessary for the “robotics challenge” constituting the last days of the event, five small-size binaural robots were equipped with adapted versions of the Two!Ears robot audition software components. The robots were placed in a specially built small-scale environment, having students solve a search-and-rescue task. In addition to advertising the software via the Two!Ears dissemination channels, this enabled direct training of external partners on the developed tools.

Two!Ears carried out a public final project event to demonstrate the work to a wider audience. As a suitable large public event that could be joined by Two!Ears, the consortium identified the “Nuit Européenne des Chercheurs” in Toulouse, on September 30th, 2016 (http://www.nuitdeschercheurs-france.eu/Toulouse). By choosing the Toulouse edition of this France-wide event, Two!Ears was able to demonstrate the

6 See http://twoears2015.sciencesconf.org and pictures from the event on Flickr https://www.flickr.com/photos/twoearsproject/.

robot deployment system in a place where this system could easily be transported to. A total of 3,400 people attended the event. Details on the event can be found on the project website. A video introduction has been provided at `https://vimeo.com/193283179`, and pictures of the event can be found at `https://www.flickr.com/photos/twoearsproject/`.

For direct dissemination of the foreground of the robotics-based head-turning-motorization and control of the KEMAR HATs, contacts with G.R.A.S. have been established. It was decided that the system would not directly be integrated into their product line. However, since Two!Ears' developments re-use all existing mechanical components of the KEMAR HATS and do not affect its general functionality, the usage of the motorization unit by others is not limited. To enable easy and open adoption, the technical documentation for manufacturing as well as the control software were made publicly available by the Two!Ears consortium.

The source code was published via Github, together representing a complete software package arranged and released via the website. It contains the stable working version of the Two!Ears model software, combined with an extensive online documentation as the user manual with examples – see `docs.twoears.eu`. The substantial documentation and general usability of the software and database – in our view highly unique for a project in the area of audio signal processing and auditory modelling – has lead to positive feedback by users (for example: "I found both the 2!Ears code and the datasets remarkably easy and convenient to use.", Antoine Deleforge, INRIA, France). The Two!Ears framework has been acknowledged as one of the candidates for the so-called binaural model challenge initiated recently in Oldenburg University which aims to compare the currently available auditory modelling tools against one another in various aspects.

List of Acronyms

Acronyms

ACF auto-correlation function

AFE auditory front-end

ALSA Advanced Linux Sound Architecture

AMS amplitude modulation spectrogram

CCF cross-correlation function

CNN convolutional neural network

DNN deep neural network

DOF degree-of-freedom

DRNL dual-resonance non-linear

EM expectation-maximization

ERB equivalent rectangular bandwidth

GMM Gaussian mixture model

IC interaural coherence

ILD interaural level difference

ITD interaural time difference

MCT multi-conditional training

MOC medial olivo-cochlear

ReLU rectified linear units

T-F time-frequency

Bibliography

(**http://spatialaudio.net/ssr** a), "The SoundScape Renderer," http://spatialaudio.net/ssr. (Cited on page 11)

(**http://www.redmine.org**), "Redmine," http://www.redmine.org. (Cited on page 40)

(**https://subversion.apache.org** a), "Apache Subversion," https://subversion.apache.org. (Cited on page 40)

(**https://git-scm.com** b), "git," https://git-scm.com. (Cited on page 40)

(**https://creativecommons.org** b), "Crative Commons," https://creativecommons.org. (Cited on page 40)

Banerjee, A., Dhillon, I. S., Ghosh, J., and Sra, S. (**2005**), "Clustering on the Unit Hypersphere Using Von Mises-Fisher Distributions," *J. Mach. Learn. Res.* **6**, pp. 1345–1382. (Cited on page 24)

Blauert, J. (**1997**), *Spatial hearing - The psychophysics of human sound localization*, The MIT Press, Cambride, MA, USA. (Cited on page 22)

Blauert, J., Braasch, J., Buchholz, J. M., Colburn, H. S., Jekosch, U., Kohlrausch, A., Mourjopoulos, J., Pulkki, V., and Raake, A. (**2009**), "Aural Assessment by Means of Binaural Algorithms - The AABBA project," in *2nd Intl. Symp. on Auditory and Audiological Research.* (Cited on page 41)

European Research Commission (), "H2020 Programme: Guidelines of FAIR Data Management in Horizon 2020," Version 3.0, 26.7.2016. (Cited on page 40)

Grondin, F., Létourneau, D., Ferland, F., Rousseau, V., and Michaud, F. (**2013**), "The ManyEars Open Framewor," *Autonomous Robots* **34**(3), pp. 217–232. (Cited on page 39)

Gustafsson, T., Rao, B. D., and Trivedi, M. (**2003**), "Source Localization in Reverberant Environments: Modeling and Statistical Analysis," *IEEE Trans. Speech and Audio Proc.* **11**(6), pp. 791–803. (Cited on page 35)

Hahn, N. and Spors, S. (**2014**a), "Measurement of Time-Variant Binaural Room Impulse Responses for Data-Based Synthesis of Dynamic Auditory Scenes," in *German Annual*

Conference on Acoustics (DAGA). (Cited on page 10)

Hahn, N. and Spors, S. (**2014**b), "Identification of Dynamic Acoustic Systems by Orthogonal Expansion of Time-Variant Impulse Responses," in *IEEE-EURASIP International Symposium on Control, Communications, and Signal Processing.* (Cited on page 10)

Hahn, N. and Spors, S. (**2015**), "Continuous Measurement of Impulse Responses on a Circle using a Uniformly Moving Microphone," in *European Signal Processing Conference.* (Cited on page 10)

Hahn, N. and Spors, S. (**2016**), "Comparison of Continuous Measurement Techniques for Spatial Room Impulse Responses," in *European Signal Processing Conference*, Budapest, Hungary, pp. 1638–1642. (Cited on page 10)

Kopco, N., Best, V., and Carlile, S. (**2010**), "Speech localization in a multitalker mixture," *The Journal of the Acoustical Society of America* **127**(3), pp. 1450–1457. (Cited on page 25)

Lecun, Y., Bottou, L., Bengio, Y., and Haffner, P. (**1998**), "Gradient-based learning applied to document recognition," *Proceedings of the IEEE* **86**(11), pp. 2278–2324. (Cited on page 24)

Liebetrau, J., Sporer, T., Kämpf, S., and Schneider, S. (**2010**), "Standardization of PEAQ-MC: Extension of ITU-R BS.1387-1 to Multichannel Audio," in *In AES 40th International Conf. Spatial Audio, Tokyo, Japan, Oct. 2010.* (Cited on page 39)

Lombard, A., Rosenkranz, T., Buchner, H., and Kellermann, W. (**2009**), "Multidimensional localization of multiple sound sources using averaged directivity patterns of blind source separation systems," in *Proc. IEEE Int. Conf. on Acoustics, Speech, and Signal Processing (ICASSP)*, Taipei, Taiwan, pp. 233–236. (Cited on page 35)

Ma, N. and Brown, G. J. (**2016**), "Speech localisation in a multitalker mixture by humans and machines," in *Proceedings of Interspeech.* (Cited on pages 25 and 36)

Ma, N., Brown, G. J., and Gonzalez, J. A. (**2015**a), "Exploiting top-down source models to improve binaural localisation of multiple sources in reverberant environments," in *Proceedings of Interspeech.* (Cited on page 24)

Ma, N., Marxer, R., Barker, J., and Brown, G. J. (**2015**b), "Exploiting synchrony spectra and deep neural networks for noise-robust automatic speech recognition," in *ASRU Workshop on the CHiME-3 Challenge.* (Cited on page 38)

Ma, N., May, T., and Brown, G. J. (**2015**c), "Robust localisation of multiple speakers exploiting deep neural networks and head movements," in *Proceedings of Interspeech*, pp. 3302–3306. (Cited on pages 21, 22, and 36)

Nakadai, K., Takahashi, T., Okuno, H., Nakajima, H., Hasegawa, Y., and Tsujino, H. (**2010**), "Design and Implementation of Robot Audition System 'HARK' Ů Open Source Software for Listening to Three Simultaneous Speakers," *Advanced Robotics* **24**(5-6), pp. 739–761. (Cited on page 39)

Raake, A. and Wierstorf, H. (**2016**), "Assessment of audio quality and experience using binaural-hearing models," in *Proceedings 22nd International Congress on Acoustics (ICA)*. (Cited on page 38)

Relano-Iborra, H., May, T., Zaar, J., Scheidiger, C., and Dau, T. (**2016**), "Predicting speech intelligibility based on a correlation metric in the envelope power spectrum domain," *Journal of the Acoustical Society of America* **140**(4), pp. 2670–2679. (Cited on page 40)

Rumsey, F. (**2010**), "Sound Quality Evaluation," *J. Audio Eng. Soc.* **58**(10), pp. 853–858. (Cited on pages 35 and 39)

Skowronek, J. and Raake, A. (**2015**), "Assessment of Cognitive Load, Speech Communication Quality and Quality of Experience for spatial and non-spatial audio conferencing calls," *Speech Communication* **66**, pp. 154–175. (Cited on page 39)

Søndergaard, P. and Majdak, P. (**2013**), "The auditory-modeling toolbox," in *The technology of binaural listening*, edited by J. Blauert, Springer, Berlin–Heidelberg–New York NY, chap. 2. (Cited on page 15)

Teki, S., Chait, M., Kumar, S., von Kriegstein, K., and Griffiths, T. (**2011**), "Brain Bases for Auditory Stimulus-Driven Figure-Ground Segregation," *Journal of Neuroscience* **31**(1), pp. 164–171. (Cited on page 37)

Teng, P., Lombard, A., and Kellermann, W. (**2010**), "Disambiguation in multidimensional tracking of multiple acoustic sources using a Gaussian likelihood criterion," in *Proc. IEEE Int. Conf. on Acoustics, Speech, and Signal Processing (ICASSP)*, Dallas, Texas, USA, pp. 145–148. (Cited on page 35)

Varga, A. and Moore, R. (**1990**), "Hidden Markov model decomposition of speech and noise," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, pp. 845–848. (Cited on page 23)

Wierstorf, H., Ende, C., and Raake, A. (**2015**), "Klangverfärbung in der Wellenfeldsynthese - Experimente und Modellierung," in *DAGA2015*. (Cited on page 38)

Winter, F., Schutz, F., and Spors, S. (**2014**), "Localization Properties of Data-based Binaural Synthesis including Translatory Head-Movements," in *Forum Acousticum*. (Cited on page 10)

Appendix I: Tables A, B

Note: Table B.1 not applicable for Two!Ears.

Table A1

List of scientific publications (peer reviewed indicated as such). Most relevant publications are indicated by bold font, based on assessment by the respective institutional representative. Note 1: Some domains such as audio technology and audiology conferences may not have peer-review but are mentioned here due to their high relevance in the field. Note 2: Invited papers are indicated as such.											
Indi-cator	Peer-review?	Title	Main author(s) and involved partners	Title of the periodical or series	Number, date or frequency	Publisher	Place of publi-cation	Year of public-cation	Relevant pages	Permanent identifiers (if available) [1]	Will this be open access? [2]
A	Yes	Predicting speech intelligibility based on a correlation metric in the envelope power spectrum domain	Relano-Iborra	Journal of the Acoustical Society of America	140(4)	Acoustical Society of America	Melville, NY	2016	2670-2679	http://dx.doi.org/10.1121/1.4964505	Yes
B	Yes	Outcome measures based on classification performance fail to predict the intelligibility of binary-masked speech	Kressner	Journal of the Acoustical Society of America	139(6)	Acoustical Society of America	Melville, NY	2016	3033-3036	http://dx.doi.org/10.1121/1.4952439	Yes

List of scientific publications (peer reviewed indicated as such). Most relevant publications are indicated by bold font, based on assessment by the respective institutional representative. Note 1: Some domains such as audio technology and audiology conferences may not have peer-review but are mentioned here due to their high relevance in the field. Note 2: Invited papers are indicated as such.											
Indi-cator	Peer-review?	Title	Main author(s) and involved partners	Title of the periodical or series	Number, date or frequency	Publisher	Place of publi-cation	Year of public-cation	Relevant pages	Permanent identifiers (if available) [1]	Will this be open access? [2]
C	No	Assessing the contribution of binaural cues for apparent source width perception via a functional model	Käsbach	Proceedings of the 22nd International Congress on Acoustics	5/9/2016	ICA	Buenos Aires, Argentina	2016	ABS-768	http://orbit.dtu.dk/en/publications/assessing-the-contribution-of-binaural-cues-for-apparent-source-width-perception-via-a-functional-model(442c7748-ab46-4fda-8d00-c47b534558e7).html	Yes
D	Yes	Introducing the Turbo-Twin-HMM for Audio-Visual Speech Enhancement	Zeiler, Kolossa	Proceedings of Interspeech 2016	8/9/2016	ISCA	San Francisco	2016		https://www.researchgate.net/publication/307889556_Introducing_the_Turbo-Twin-HMM_for_Audio-Visual_Speech_Enhancement	Yes

List of scientific publications (peer reviewed indicated as such). Most relevant publications are indicated by bold font, based on assessment by the respective institutional representative. Note 1: Some domains such as audio technology and audiology conferences may not have peer-review but are mentioned here due to their high relevance in the field. Note 2: Invited papers are indicated as such.											
Indi-cator	Peer-review?	Title	Main author(s) and involved partners	Title of the periodical or series	Number, date or frequency	Publisher	Place of publi-cation	Year of public-cation	Relevant pages	Permanent identifiers (if available) [1]	Will this be open access? [2]
E	Yes	Speech localisation in a multitalker mixture by humans and machines	Ma, Brown (USFD)	Proceedings of Interspeech 2016	8/9/2016	ISCA	San Francisco	2016	1149-1152	Not yet available	Yes
F	Yes	A robust dual-microphone speech source localization algorithm for reverberant environments	Guo, Ma, Brown (USFD)	Proceedings of Interspeech 2016	8/9/2016	ISCA	San Francisco	2016	1063-1066	http://eprints.whiterose.ac.uk/102624/	Yes
G	Yes	Comparing the influence of spectro-temporal integration in computational speech segregation	Bentsen	Proceedings of Interspeech 2016	8/9/2016	ISCA	San Francisco	2016			Yes
H	No	Active Localization of Sound Sources using Binaural Models	Schymura	Proceedings of DAGA		DEGA	Aachen	2016			Yes
I	Yes	Robust audiovisual speech recognition using noise-adaptive linear discriminant analysis	Zeiler	Proceedings of ICASSP		IEEE		2016			Yes

List of scientific publications (peer reviewed indicated as such). Most relevant publications are indicated by bold font, based on assessment by the respective institutional representative. Note 1: Some domains such as audio technology and audiology conferences may not have peer-review but are mentioned here due to their high relevance in the field. Note 2: Invited papers are indicated as such.											
Indi-cator	Peer-review?	Title	Main author(s) and involved partners	Title of the periodical or series	Number, date or frequency	Publisher	Place of publi-cation	Year of public-cation	Relevant pages	Permanent identifiers (if available) [1]	Will this be open access? [2]
J	Yes	Uncertain LDA: Including observation uncertainties in discriminative transforms	Saeidi	IEEE Trans. on Pattern Analysis and Machine Intelligence		IEEE		2015		https://doi.org/10.1109/TPAMI.2015.2481420	No
K	Yes	Exploiting synchrony spectra and deep neural networks for noise-robust automatic speech recognition	Ma, Brown (USFD)	ASRU Workshop on the CHiME-3 Challenge		ASRU	Scottsdale, Arizona	2015		http://ieeexplore.ieee.org/document/7404835/	Yes
L	Yes	The role of temporal resolution in modulation-based speech segregation	May	Proceedings of Interspeech 2015		ISCA		2015	170-174		Yes
M	Yes	**Binaural Sound Source Localisation and Tracking using a Dynamic Spherical Head Model**	Schymura	Proceedings of Interspeech		ISCA		2015		http://www.isca-speech.org/archive/interspeech_2015/i15_0165.html	Yes

List of scientific publications (peer reviewed indicated as such). Most relevant publications are indicated by bold font, based on assessment by the respective institutional representative. Note 1: Some domains such as audio technology and audiology conferences may not have peer-review but are mentioned here due to their high relevance in the field. Note 2: Invited papers are indicated as such.

Indi-cator	Peer-review?	Title	Main author(s) and involved partners	Title of the periodical or series	Number, date or frequency	Publisher	Place of publi-cation	Year of public-cation	Relevant pages	Permanent identifiers (if available) [1]	Will this be open access? [2]
N	Yes	**Exploiting top-down source models to improve binaural localisation of multiple sources in reverberant environments**	Ma, Brown (USFD)	Proceedings of Interspeech		ISCA		2015		http://eprints.whiterose.ac.uk/102627/	Yes
O	Yes	**Exploiting deep neural networks and head movements for binaural localisation of multiple speakers in reverberant conditions**	Ma, Brown, May (USFD, DTU)	Proceedings of Interspeech		ISCA		2015		http://eprints.whiterose.ac.uk/102628/	Yes
P	Yes	Robust localisation of multiple speakers exploiting head movements and multi-conditional training of binaural cues	May, Ma, Brown (DTU, USFD)	Proceedings of IEEE ICASSP		IEEE		2015	2679-2683	http://eprints.whiterose.ac.uk/90481/	Yes

List of scientific publications (peer reviewed indicated as such). Most relevant publications are indicated by bold font, based on assessment by the respective institutional representative. Note 1: Some domains such as audio technology and audiology conferences may not have peer-review but are mentioned here due to their high relevance in the field. Note 2: Invited papers are indicated as such.											
Indi-cator	Peer-review?	Title	Main author(s) and involved partners	Title of the periodical or series	Number, date or frequency	Publisher	Place of publi-cation	Year of public-cation	Relevant pages	Permanent identifiers (if available) [1]	Will this be open access? [2]
Q	Yes	A machine-hearing system exploiting head movements for binaural sound localisation in reverberant conditions	Ma, May, Wierstorf, Brown (USFD,DTU,TUB)	Proceedings of IEEE ICASSP		IEEE		2015	2699-2703	http://ieeexplore.ieee.org/document/7178461/	Yes
R	Yes	Textured Object Recognition: Balancing Model Robustness and Complexity	Manfredi	16th Int. Conf. on Computer Analysis of Images and Patterns (CAIP'2015)				2015			Yes
S	No	The effect of interaural-time-difference fluctuations on apparent source width	Käsbach	Proc. 7th Forum Acusticum			Krakow. Poland	2014			Yes
T	Yes	Computational speech segregation based on an auditory-inspired modulation analysis	May	Journal of the Acoustical Society of America	136(6)	Acoustical Society of America	Melville, NY	2015	3350-3359	http://dx.doi.org/10.1121/1.4901711	Yes
U	Yes	Requirements for the evaluation of computational speech segregation systems	May	Journal of the Acoustical Society of America	136(6)	Acoustical Society of America	Melville, NY	2014	EL398-EL404	http://dx.doi.org/10.1121/1.4901133	Yes

List of scientific publications (peer reviewed indicated as such). Most relevant publications are indicated by bold font, based on assessment by the respective institutional representative. Note 1: Some domains such as audio technology and audiology conferences may not have peer-review but are mentioned here due to their high relevance in the field. Note 2: Invited papers are indicated as such.											
Indi-cator	Peer-review?	Title	Main author(s) and involved partners	Title of the periodical or series	Number, date or frequency	Publisher	Place of publi-cation	Year of public-cation	Relevant pages	Permanent identifiers (if available) [1]	Will this be open access? [2]
V	Yes	Generalization of supervised learning for binary mask estimation	May	14th International Workshop on Acoustic Signal Enhancement (IWAENC)				2014	154-158		Yes
W	Yes	Comparing human and automatic speech recognition in a perceptual restoration experiment	Remes, Brown (USFD)	Computer Speech and Language	35	Elsevier	Netherlands	2016	14-31	http://dx.doi.org/10.1016/j.csl.2015.06.005	Yes
X	Yes	Feature enhancement of reverberant speech by distribution matching and non-negative matrix factorization	Keronen, Brown (USFD)	EURASIP Journal on Advances in Signal Processing	1	Springer	Berlin	2015	76-86	http://dx.doi.org/10.1186/s13634-015-0259-1	Yes
Y	No	**Binaural sound source localisation using a Bayesian-network-based blackboard system and hypothesis-driven feedback**	Schymura, Walther, Kolossa, Ma, Brown (RUB,USFD)	Proc.7th Forum Acousticum	2014		Poland	2014		http://eprints.whiterose.ac.uk/87104/	Yes

List of scientific publications (peer reviewed indicated as such). Most relevant publications are indicated by bold font, based on assessment by the respective institutional representative. Note 1: Some domains such as audio technology and audiology conferences may not have peer-review but are mentioned here due to their high relevance in the field. Note 2: Invited papers are indicated as such.											
Indi-cator	Peer-review?	Title	Main author(s) and involved partners	Title of the periodical or series	Number, date or frequency	Publisher	Place of publi-cation	Year of public-cation	Relevant pages	Permanent identifiers (if available) [1]	Will this be open access? [2]
Z	**Yes**	**Robust Detection of Environmental Sounds in Binaural Auditory Scenes**	Trowitzsch	Transactions on Audio, Speech and Language Processing	2016	IEEE (under review)	N/A	2016	N/A	N/A	Yes
AA	No	Aufmerksam hören (Attentive listening)	Blauert, Walther	DAGA 2017 abstract available	2017		Kiel	2017 to come			Yes
AB	Yes	Komplexe instrumentelle Sound-Qualitätsbe-urteilung als Ausgangspunkt für Schätzungen des Kulturgrades der Rezipienten – ein Versuch (Complex instrumental judgements on sound quality as a starting point for estimations of the cultural level	Blauert, Raake,	Schmidt, W.G. (Ed.) Die Natur-Kultur-Grenze in Kunst und Wissenschaft	2014	K & N Verlag Königshausen und Neumann	Würzburg, Germany		193–214		Yes
AC	No	**Psychoakustik aus perzeptionistischer Sicht (Psychoacoustics from a perceptualist's point of view)**	Blauert	Proc. 18th Jahrestg. Dtsch. Ges. Audiol., DGA (invited plenary keynote lecture)	2016		Bochum, Germany	2016			Yes

List of scientific publications (peer reviewed indicated as such). Most relevant publications are indicated by bold font, based on assessment by the respective institutional representative. Note 1: Some domains such as audio technology and audiology conferences may not have peer-review but are mentioned here due to their high relevance in the field. Note 2: Invited papers are indicated as such.											
Indi-cator	Peer-review?	Title	Main author(s) and involved partners	Title of the periodical or series	Number, date or frequency	Publisher	Place of publi-cation	Year of public-cation	Relevant pages	Permanent identifiers (if available) [1]	Will this be open access? [2]
AD	No	The advent of Communication Acoustics in retrospect	Blauert	Proc. Intl. Congr. Acoustics, ICA 2016, (invited keynote paper) (also submitted to POMA)	2016		Buenos Aires, Argentina	2016	paper #187		Yes
AE	Yes	**Auditory perception in rooms**	Braasch, Blauert	N. Xiang (ed). Architectural Acoustics Handbook	2016	J. Ross Publishing, Inc.	Fort Lauderdale, FL, USA	Jan. 2017	173–196		Yes
AF	Yes	**A Precedence-effect model with top-down processing stages based on visual cues**	Braasch, Pastore, Deshpande, Blauert	Proc. Intl. Congr. Acoustics, ICA 2016, (invited paper) (also submitted to POMA)	2016		Buenos Aires, Argentina	2016	paper #827		Yes
AG	No	**A bi-modal model to simulate auditory expectation for reverberation time and direct-to-reverberant energy from visual feedback**	Braasch, Pastore, Deshpande, Blauert	169^{th} Meeting Acoust. Soc. Am. (invited talk)	2015		Pittsburg PA, USA	2015			Yes

List of scientific publications (peer reviewed indicated as such). Most relevant publications are indicated by bold font, based on assessment by the respective institutional representative. Note 1: Some domains such as audio technology and audiology conferences may not have peer-review but are mentioned here due to their high relevance in the field. Note 2: Invited papers are indicated as such.											
Indi-cator	Peer-review?	Title	Main author(s) and involved partners	Title of the periodical or series	Number, date or frequency	Publisher	Place of publi-cation	Year of public-cation	Relevant pages	Permanent identifiers (if available) [1]	Will this be open access? [2]
AH	No	A case for Two!Ears in audio-quality assessment	Raake, Wierstorf, Blauert	Proc. 7th Forum Acusticum (invited paper),	2014		Krakow, Poland	2014	Paper SS16-19		Yes
AI	No	System zur Simulation von kognitivem Feedback im Kontext auditiver Szenenanalyse und auditiver Qualitätsbeurteilung (A system for the simulation of cognitive feedback in the context of auditory scene analysis and auditory sound-quality assessment).	Walther, Blauert, Raake	Fortschr. Akustik, DAGA 2016, Dtsch. Ges. Akustik	2016		Oldenburg, Germany	2016	584–585		Yes
AJ	Yes	**Simulating cognitive feedback in the context of binaural scene analysis (also submitted to POMA)**	Walther, Blauert	Proc. Intl. Congr. Acoustics, ICA 2016, (invited paper) (also submitted to POMA)	2016		Buenos Aires, Argentina	2016	paper #154		Yes

List of scientific publications (peer reviewed indicated as such). Most relevant publications are indicated by bold font, based on assessment by the respective institutional representative. Note 1: Some domains such as audio technology and audiology conferences may not have peer-review but are mentioned here due to their high relevance in the field. Note 2: Invited papers are indicated as such.											
Indi-cator	Peer-review?	Title	Main author(s) and involved partners	Title of the periodical or series	Number, date or frequency	Publisher	Place of publi-cation	Year of public-cation	Relevant pages	Permanent identifiers (if available) [1]	Will this be open access? [2]
AK	Yes	The influence of signal type on the internal auditory representation of a room	Teret, Pastore, Braasch	Journal of the Acoustical Society of America	2016	Acoustical Society of America	Melville, NY, USA	N/A	N/A	http://dx.doi.org/10.1121/1.4920478	Yes
AL	Yes	**Source-blind binaural source segregation utilizing head movement**	Desphande, Braasch	Proc. Intl. Congr. Acoustics, ICA 2016 (invited paper)	2016		Buenos Aires, Argentina	2016	paper #818		Yes
AM	Yes	**A binaural model to segregate sound sources in the presence of early reflections using a multi-source precedence-effect model**	Braasch, Desphande	Proc. Intl. Congr. Acoustics, ICA 2016 (invited paper)	2016		Buenos Aires, Argentina	2016	Paper #43		Yes
AN	**Yes**	**Blind localization and segregation of two sources including a binaural head movement model**	Desphande, Braasch	Journal of the Acoustical Society of America/ Express Letter	N/A	Acoustical Society of America	Melville, NY	2016, submitted	N/A		Yes

List of scientific publications (peer reviewed indicated as such). Most relevant publications are indicated by bold font, based on assessment by the respective institutional representative. Note 1: Some domains such as audio technology and audiology conferences may not have peer-review but are mentioned here due to their high relevance in the field. Note 2: Invited papers are indicated as such.											
Indi-cator	Peer-review?	Title	Main author(s) and involved partners	Title of the periodical or series	Number, date or frequency	Publisher	Place of publi-cation	Year of public-cation	Relevant pages	Permanent identifiers (if available) [1]	Will this be open access? [2]
AO	Yes	Monte Carlo Exploration for Active Binaural Localization	Schymura	ICASSP (accepted for publication)	2017	IEEE	New Orleans, LA, USA	2017			Yes
AP	Yes	Improving Audio-Visual Speech Recognition using Deep Neural Networks with Dynamic Stream Reliability Estimates	Meutzner	ICASSP (accepted for publication)	2017	IEEE	New Orleans, LA, USA	2017			Yes
AQ	Yes	**Modulating the Auditory Turn-to Reflex on the Basis of Multimodal Feedback Loops: the Dynamic Weighting Model**	Cohen-Lhyver	ROBIO	2015	IEEE	Zhuhai, China	2015			Yes

List of scientific publications (peer reviewed indicated as such). Most relevant publications are indicated by bold font, based on assessment by the respective institutional representative. Note 1: Some domains such as audio technology and audiology conferences may not have peer-review but are mentioned here due to their high relevance in the field. Note 2: Invited papers are indicated as such.											
Indi-cator	Peer-review?	Title	Main author(s) and involved partners	Title of the periodical or series	Number, date or frequency	Publisher	Place of publi-cation	Year of public-cation	Relevant pages	Permanent identifiers (if available) [1]	Will this be open access? [2]
AR	Yes	**Multimodal fusion and inference using binaural audition and vision**	Cohen-Lhyver	Intl. Congr. Acoustics ICA 2016 (invited paper)	2016		Buenos Aires, Argentina	2016			Yes
AS	Yes	**An Information-based Feedback Control for Audio-Motor Binaural Localization**	Bustamante, G. and Danès, P. and Forgue, T. and Podlubne, A. and Manhès, J.	Autonomous Robots (Journal) - Special Issue on Active Perception (under second review)		Springer					**Yes**
AT	Yes	A One-step-ahead Information- based Feedback Control for Binaural Active Localization	Bustamante, G. and Danès, P. and Forgue, T. and Podlubne, A.	European Signal Processing Conference			Budapest, Hungary	2016		www.eurasip.org/Proceedings/Eusipco/Eusipco2016/papers/1570256309.pdf	Yes

List of scientific publications (peer reviewed indicated as such). Most relevant publications are indicated by bold font, based on assessment by the respective institutional representative. Note 1: Some domains such as audio technology and audiology conferences may not have peer-review but are mentioned here due to their high relevance in the field. Note 2: Invited papers are indicated as such.											
Indi-cator	Peer-review?	Title	Main author(s) and involved partners	Title of the periodical or series	Number, date or frequency	Publisher	Place of publi-cation	Year of public-cation	Relevant pages	Permanent identifiers (if available) [1]	Will this be open access? [2]
AU	Yes	Towards Information-Based Feedback Control for Binaural Active Localization	Bustamante, G. and Danès, P. and Forgue, T. and Podlubne, A.	IEEE Int. Conf. on Acoustics, Speech and Signal Processing		IEEE	Shanghai, China	2016		http://ieeexplore.ieee.org/document/7472894/	Yes
AV	Yes	A Three-Stage Framework to Active Source Localization from a Binaural Head	Bustamante, G. and Portello, A. and Danès, P.	IEEE Int. Conf. on Acoustics, Speech and Signal Processing		IEEE	Brisbane, Australia	2015		http://ieeexplore.ieee.org/document/7179047/	Yes
AW	No	Active Localization of an Intermittent Sound Source from a Moving Binaural Sensor	Portello, A. and Bustamante, G. and Danès, P. and Piat, J. and Manhès, J.	Proc. 7th Forum. 7th Forum Acusticum			Krakow, Krakow, Poland	2014			Yes
AX	Yes	Positioning of Musical Foreground Parts in Surrounding Sound Stages	Hold, Nagel, Wierstorf, Raake	AES Int. Conf. on Audio for Virtual and Augmented Reality		AES	Los Angeles, USA	2016	Paper 7-2		Yes
AY	No	Assessment of audio quality and experience using binaural-hearing models	Raake, Wierstorf	Proc. Intl. Congr. Acoustics, ICA 2016 (invited paper)			Buenos Aires, Argentina	2016			Yes

List of scientific publications (peer reviewed indicated as such). Most relevant publications are indicated by bold font, based on assessment by the respective institutional representative. Note 1: Some domains such as audio technology and audiology conferences may not have peer-review but are mentioned here due to their high relevance in the field. Note 2: Invited papers are indicated as such.											
Indi-cator	Peer-review?	Title	Main author(s) and involved partners	Title of the periodical or series	Number, date or frequency	Publisher	Place of publi-cation	Year of public-cation	Relevant pages	Permanent identifiers (if available) [1]	Will this be open access? [2]
AZ	No	The Difference Between Stereophony and Wave Field Synthesis in the Context of Popular Music	Hold, Wierstorf, Raake	140th Conv. AES		AES	Paris, France	2016	Paper 9533		Yes
BA	No	Perceptual assessment of spatial sound: the Two!Ears project	Wierstorf	140th Conv. AES		AES	Paris, France	2016	Invited talk		Yes
BB	No	Tonmischung für Stereophonie und Wellenfeldsynthese im Vergleich	Hold, Wierstorf, Raake	Fortschr. Akustik, DAGA 2016, Dtsch. Ges. Akustik		DEGA	Aachen, Germany	2016	1023-1026		Yes
BC	No	Auf dem Weg zu binauraler Modellierung mit Kognition: das Two!Ears Modell	Wierstorf, Raake	Fortschr. Akustik, DAGA 2016, Dtsch. Ges. Akustik		DEGA	Aachen, Germany	2016	Presenta tion only		Yes
BD	No	Klangverfärbung in der Wellenfeldsynthese - Experimente und Modellierung	Wierstorf, Ende, Raake	Fortschr. Akustik, DAGA 2015, Dtsch. Ges. Akustik		DEGA	Nurembe rg, Germany	2015	490-493		Yes

List of scientific publications (peer reviewed indicated as such). Most relevant publications are indicated by bold font, based on assessment by the respective institutional representative. Note 1: Some domains such as audio technology and audiology conferences may not have peer-review but are mentioned here due to their high relevance in the field. Note 2: Invited papers are indicated as such.

Indi-cator	Peer-review?	Title	Main author(s) and involved partners	Title of the periodical or series	Number, date or frequency	Publisher	Place of publi-cation	Year of public-cation	Relevant pages	Permanent identifiers (if available) [1]	Will this be open access? [2]
BE	No	Audioqualitätsbeurteilung: Ein Fall für TWO!EARS	Raake, Wierstorf, Blauert	Fortschr. Akustik, DAGA 2015, Dtsch. Ges. Akustik		DEGA	Nuremberg, Germany	2015			Yes
BF	No	Predicting localization accuracy for stereophonic downmixes in Wave Field Synthesis	Wierstorf, Spors	Proc. Forum Acusticum		EAA	Krakow, Poland	2014			Yes
BG	Yes	Assessing localization accuracy in sound field synthesis	Wierstorf, Raake, Spors	Journal of the Acoustical Society of America		Acoustical Society of America	Melville, NY, USA	2017	Accepted		Yes
BH	**Yes**	**On Analytic Methods for 2.5-D Local Sound Field Synthesis Using Circular Distributions of Secondary Sources**	Winter, F.; Ahrens, J.; Spors, S.	Transactions on Audio, Speech, and Language Processing,	vol. 24, no. 5	IEEE/ACM		2016	914-926	https://doi.org/10.1109/TASLP.2016.2531902	No
BI	**Yes**	**On Fractional Delay Interpolation for Local Wave Field Synthesis**	Winter, F.; Spors, S.	European Signal Processing Conference	24th	EURASIP, IEEE	Budapest, Hungary	2016	2415-2419	https://doi.org/10.1109/EUSIPCO.2016.7760682	Yes

List of scientific publications (peer reviewed indicated as such). Most relevant publications are indicated by bold font, based on assessment by the respective institutional representative. Note 1: Some domains such as audio technology and audiology conferences may not have peer-review but are mentioned here due to their high relevance in the field. Note 2: Invited papers are indicated as such.											
Indi-cator	Peer-review?	Title	Main author(s) and involved partners	Title of the periodical or series	Number, date or frequency	Publisher	Place of publi-cation	Year of public-cation	Relevant pages	Permanent identifiers (if available) [1]	Will this be open access? [2]
BJ	Yes	Comparison of Continuous Measurement Techniques for Spatial Room Impulse Responses	Hahn, N.; Spors, S.:	European Signal Processing Conference	24th	EURASIP, IEEE	Budapest, Hungary	2016	1638-1642	https://doi.org/10.1109/EUSIPCO.2016.7760526	Yes
BK	No	Local Wave Field Synthesis by Spatial Band-limitation in the Circular/Spherical Harmonics Domain	Hahn, N.; Winter, F.; Spors, S.	140th Convention of the Audio Engineering Society		AES	Paris, France	2016			Yes
BL	**No**	**Database of Binaural Room Impulse Responses of an Apartment-Like Environment**	Winter, F.; Wierstorf, H.; Podlubne, A.; Forgue, T.; Manhès, J.; Herrb, M.; Spors, S.; Raake, A.; Danès, P.	140th Convention of the Audio Engineering Society		AES	Paris, France	2016			Yes
BM	No	A Comparison of Sound Field Synthesis Techniques for Non-Smooth Secondary Source Distributions	Winter, F.; Spors, S..	German Annual Conference on Acoustics (DAGA)		DEGA	Aachen, Germany	2016	1463-1466		Yes

List of scientific publications (peer reviewed indicated as such). Most relevant publications are indicated by bold font, based on assessment by the respective institutional representative. Note 1: Some domains such as audio technology and audiology conferences may not have peer-review but are mentioned here due to their high relevance in the field. Note 2: Invited papers are indicated as such.											
Indi-cator	Peer-review?	Title	Main author(s) and involved partners	Title of the periodical or series	Number, date or frequency	Publisher	Place of publi-cation	Year of public-cation	Relevant pages	Permanent identifiers (if available) [1]	Will this be open access? [2]
BN	No	Analysis of time-varying system identification using the Normalized Least Mean Square algorithm in the context of data-based binaural synthesis	Hahn, N.; Spors, S.:	German Annual Conference on Acoustics (DAGA)		DEGA	Aachen, Germany	2016	1012-1015		Yes
BO	Yes	Binaural Sound Source Localisation and Tracking using a dynamic Spherical Head Model	Schymura, C.; Winter, F.; Kolossa, D.; Spors, S.	Interspeech			Dresden, Germany	2015			Yes
BP	No	Physical Properties of Local Wave Field Synthesis using Circular Loudspeaker Arrays	Winter, F.; Spors, S.	EuroNoise			Maastricht, The Netherlands	2015			Yes
BQ	Yes	Physical Properties of Local Wave Field Synthesis using Linear Loudspeaker Arrays	Winter, F.; Spors, S.:	138th Convention of the Audio Engineering Society			Warsaw, Poland	2015			Yes

List of scientific publications (peer reviewed indicated as such). Most relevant publications are indicated by bold font, based on assessment by the respective institutional representative. Note 1: Some domains such as audio technology and audiology conferences may not have peer-review but are mentioned here due to their high relevance in the field. Note 2: Invited papers are indicated as such.											
Indi-cator	Peer-review?	Title	Main author(s) and involved partners	Title of the periodical or series	Number, date or frequency	Publisher	Place of publi-cation	Year of public-cation	Relevant pages	Permanent identifiers (if available) [1]	Will this be open access? [2]
BR	No	Parameter Analysis for Range Extrapolation of Head-Related Transfer Functions using Virtual Local Wave Field Synthesis	Winter, F.; Spors, S.:	German Annual Conference on Acoustics (DAGA)			Nurembe rg, Germany	2015			Yes
BS	No	Localization Properties of Data-based Binaural Synthesis including Translatory Head-Movements	Winter, F.; Schultz, F.; Spors, S.	7th Forum Acusticum			Krakow, Poland	2014			Yes
BT	Yes	**Can current room-acoustics indices specify the quality of experience in concert halls?**	Blauert, Raake	Psychomusicology , music, mind, and brain 25				2015			Yes

[1] A permanent identifier should be a persistent link to the published version full text if open access or abstract if article is pay per view) or to the final manuscript accepted for publication (link to article in repository).
[2] Open Access is defined as free of charge access for anyone via Internet. Please answer "yes" if the open access to the publication is already established and also if the embargo period for open access is not yet over but you intend to establish open access afterwards.

Table A2

List of dissemination activities								
No	Type of activity[1]	Main leader	Title	Date/period	Place	Type of audience[2]	Size of audience	Countries addressed
1	Web	Raake	TWO!EARS website (www.twoears.eu)	01/2014	Web	All types	Worldwide	Worldwide
2	Web	Raake	Twitter channel	01/2014	Web	All types	Worldwide	Worldwide
3	Press	Raake	TUB “Wie hört der Mensch?”	07/03/2014	TUB intern / http://www.tu-berlin.de/?id=145434	All types	Germany / TU Berlin	Germany and possibly other German-speaking countries
4	Press	Spors	“Wie hört der Mensch”	25/03/2014	http://www.uni-rostock.de/detailseite/news-artikel/wie-hoert-der-mensch/	All types	Germany	Germany and possibly other German-speaking countries
5	Interview	Spors	NDR Info – Logo – Das Wissenschaftsmagazin	02/05/2014	Public radio, Germany	Civil Society	Germany	Germany and possibly other German-speaking countries
6	Press	Kohlrausch	Presseinformation 86, Das menschliche Hören mit Robotern erforschen	30/05/2014	RUB / http://aktuell.ruhr-uni-bochum.de/pm2014/pm00086.html.de	All types	Germany / RUB	Germany and possibly other German-speaking countries
7	articles published in the popular press	Blauert, Kolossa	“Roboter soll die Welt verstehen”	23/07/2014	Westdeutsche Allgemeine Zeitung (WAZ)	Civil Society	ca. 50.000	Germany

List of dissemination activities								
No	Type of activity[1]	Main leader	Title	Date/period	Place	Type of audience[2]	Size of audience	Countries addressed
8	videos	Blauert, Kolossa	“Bochumer Forscher bringen Robos das Hören bei”	25/08/2014	Bild-Zeitung Online	Civil Society	Germany	Germany
9	articles published in the popular press	Blauert, Kolossa	“Bochumer Forscher bringen Robos das Hören bei”	25/08/2014	Bild-Zeitung	Civil Society	ca. 100.000	Germany
10	Press	Raake	“WIllst du wohl hören?” - fif-forum 2014 (www.fif-foroum2014.de)	04/12/2014	Article published in popular press / industry-press	Industry	Germany / Germanophone	Germany / Germanophone
11	Invited Seminar	Ma	Machine hearing exploiting head movements for binaural sound localisation in reverberant conditions	31/03/2015	MRC Institute of Hearing Research, Nottingham	Scientific community	30	UK
12	Other	Decorsière	Two!Ears Auditory Front-end 1.0	14/08/2015	Zonedo	Scientific community, Industry	Worldwide	Worldwide
13	Other	Winter	Two!Ears Binaural Simulator 1.0	14/08/2015	Zonedo	Scientific community, Industry	Worldwide	Worldwide
14	Invited Seminar	Ma	Exploiting top-down source models to improve binaural sound localisation	30/09/2015	Institute of Acoustics, Chinese Academy of Sciences, Beijing	Scientific community	40	China
15	Invited seminar	Danès	‘Hw/Sw Integration and Sensorimotor Feedback in Two!Ears’ at a EARS EU Project Meeting	26/11/2015	SoftBank Headquarters, Paris	EARS EU Project	~15	EARS consortium

List of dissemination activities								
No	Type of activity[1]	Main leader	Title	Date/period	Place	Type of audience[2]	Size of audience	Countries addressed
16	Invited talk	Braasch, Pastore, Deshpande, Blauert	A bi-modal model to simulate auditory expectation for reverberation time and direct-to-reverberant energy from visual feedback	2015	169th Meeting Acoust. Soc. Am.	Acousticians in all fields	ca. 50 worldwide	Pittsburg PA, USA
17	Invited plenary keynote lecture	Blauert	Psychoakustik aus perzeptionistischer Sicht (Psychoacoustics from a perceptualist’s point of view)	2016	Proc. 18th Jahrestg. Dtsch. Ges. Audiol., DGA	Acousticians in all fields	ca. 1000 Germany	Bochum, Germany
18	Invited talk	Wierstorf	Perceptual assessment of spatial sound: the Two!Ears project	2016	140th Conv. AES	Acousticians in all fields	ca. 100 worldwide	Paris, France
19	Other	Wierstorf	Sound Field Synthesis Toolbox 2.1.0	10/03/2016	Zonedo	Scientific community, Industry	Worldwide	Worldwide
20	Other	Winter	Binaural room impulse responses of an apartment-like environment	08/04/2016	Zonedo	Scientific community, Industry	Worldwide	Worldwide
21	Other	Wierstorf	Binaural room impulse responses of a 5.0 surround setup for different listening positions	13/04/2016	Zonedo	Scientific community, Industry	Worldwide	Worldwide

List of dissemination activities								
No	Type of activity[1]	Main leader	Title	Date/period	Place	Type of audience[2]	Size of audience	Countries addressed
22	Presentation	Kolossa	Listening strategies: Bayesian and active approaches to environmental robustness	29/05/2016	Workshop “Sensing: From minds to machines” at Ben Gurion University, Be’er Sheva, Israel	Scientific community	100	Israel
23	Other	Wierstorf	Supplementary material for the publication "Assessing localization accuracy in sound field synthesis"	04/06/2016	Zonedo	Scientific community, Industry	Worldwide	Worldwide
24	Other	Wierstorf	Binaural room scanning files for sound field synthesis localization experiment	13/06/2016	Zonedo	Scientific community, Industry	Worldwide	Worldwide
25	Other	Wierstorf	Head-related impulse responses of a loudspeaker array	13/06/2016	Zonedo	Scientific community, Industry	Worldwide	Worldwide
26	Other	Hold	Recordings for loudness analysis of the music mixes for comparison of wave field synthesis, surround, and stereo	13/06/2016	Zonedo	Scientific community, Industry	Worldwide	Worldwide
27	Other	Wierstorf	Binaural recordings of a real and binaural simulated circular loudspeaker array	13/06/2016	Zonedo	Scientific community, Industry	Worldwide	Worldwide
28	Other	Wierstorf	Head-related impulse responses of a loudspeaker array	13/06/2016	Zonedo	Scientific community, Industry	Worldwide	Worldwide

List of dissemination activities								
No	Type of activity[1]	Main leader	Title	Date/period	Place	Type of audience[2]	Size of audience	Countries addressed
29	Other	Wierstorf	A Free Database of Head-Related Impulse Response Measurements in the Horizontal Plane with Multiple Distances	13/06/2016	Zonedo	Scientific community, Industry	Worldwide	Worldwide
30	Other	Wierstorf	Binaural room scanning files for a 56-channel circular loudspeaker array	13/06/2016	Zonedo	Scientific community, Industry	Worldwide	Worldwide
31	Other	Hold	Signal feeds for creating the music mixes for comparison of wave field synthesis, surround, and stereo	15/06/2016	Zonedo	Scientific community, Industry	Worldwide	Worldwide
32	Other	Hold	Variations of pop mixes for Wave Field Synthesis	24/08/2016	Zonedo	Scientific community, Industry	Worldwide	Worldwide
33	Other	Hold	Object-based audio scene files for variations of the spatial arrangement in pop mixes for Wave Field Synthesis	28/08/2016	Zonedo	Scientific community, Industry	Worldwide	Worldwide
34	Invited Seminar	Blauert	International Conference on acoustics	9/2016	Buenos Aires	Scientific community	120	Worldwide
34	Invited Seminar	Raake	International Conference on Acoustics	9/2016	Buenos Aires	Scientific Community	120	Worldwide

List of dissemination activities								
No	Type of activity[1]	Main leader	Title	Date/period	Place	Type of audience[2]	Size of audience	Countries addressed
35	Exhibition	Danès	Nuit Européenne des Chercheurs	30/9/2016	Toulouse	Civil Society	3400	France
36	Invited talk	Forgue	‘Human audition, model for the robots’ at ‘Journée Science et Musique’	1/10/2016	Rennes	Civil Society	~100	France
37	Other	Wierstorf	Binaural room impulse responses recorded with KEMAR in a mid-size lecture hal	14/10/2016	Zonedo	Scientific community, Industry	Worldwide	Worldwide
38	Other	Wierstorf	Binaural room impulse responses recorded with KEMAR in a small meeting room	14/10/2016	Zonedo	Scientific community, Industry	Worldwide	Worldwide
39	Other	Wierstorf	Binaural room impulse responses recorded with KEMAR of a 19-channel linear loudspeaker array	14/10/2016	Zonedo	Scientific community, Industry	Worldwide	Worldwide
40	Other	Wierstorf	Binaural room impulse responses of a 5.0 surround setup for different listening positions	14/10/2016	Zonedo	Scientific community, Industry	Worldwide	Worldwide
41	Other	Wierstorf	Listening preferences for variations of pop mixes in Wave Field Synthesis	20/10/2016	Zonedo	Scientific community, Industry	Worldwide	Worldwide

List of dissemination activities								
No	Type of activity[1]	Main leader	Title	Date/period	Place	Type of audience[2]	Size of audience	Countries addressed
42	Invited Seminar	Ma	Exploiting deep neural networks for robust binaural localisation of multiple sources in reverberant environments	26/10/2016	Department of Electronic Engineering, University of York	Scientific community	20	UK
43	Other	Wierstorf	Listening preferences for the different reproduction systems Stereo, Surround, and Wave Field Synthesis in the context of popular music	02/11/2016	Zonedo	Scientific community, Industry	Worldwide	Worldwide
44	Other	Wierstorf	Listening position preference for different 5.0 reproductions -- data	02/11/2016	Zonedo	Scientific community, Industry	Worldwide	Worldwide
45	Other	Wierstorf	Scene related sound quality -- data	03/11/2016	Zonedo	Scientific community, Industry	Worldwide	Worldwide
46	Other	Wierstorf	Coloration of a point source in Wave Field Synthesis revisited -- data	03/11/2016	Zonedo	Scientific community, Industry	Worldwide	Worldwide
47	Other	Wierstorf	Listening test results for sound field synthesis localization experiment -- head movement data	03/11/2016	Zonedo	Scientific community, Industry	Worldwide	Worldwide
48	Other	Wierstorf	Coloration of a point source in Wave Field Synthesis -- data	03/11/2016	Zonedo	Scientific community, Industry	Worldwide	Worldwide

List of dissemination activities								
No	Type of activity[1]	Main leader	Title	Date/period	Place	Type of audience[2]	Size of audience	Countries addressed
49	Other	Wierstorf	Localisation of a real vs. binaural simulated point source -- data	03/11/2016	Zonedo	Scientific community, Industry	Worldwide	Worldwide
50	Presentation	Brown	Speech localisation in a multitalker mixture by humans and machines	04/11/2016	Hearing4All International Symposium, Oldenburg	Scientific community	200	Europe and USA
51	Other	Wierstorf	Code to reproduce the figures in the paper 'Assessing localization accuracy in sound field synthesis'	15/11/2016	Zonedo	Scientific community, Industry	Worldwide	Worldwide
52	Other	Wierstorf	Listening preferences for variations of pop mixes in a circular Wave Field Synthesis system	17/11/2016	Zonedo	Scientific community, Industry	Worldwide	Worldwide
53	Video	Brown	Two!Ears demonstration	30/11/2016	https://vimeo.com/193283179	Civil Society	Worldwide	Worldwide
54	Other	Trowitzsch	Auditory Machine Learning Training and Testing Pipeline: AMLTTP v2.0-alpha	12/2016	Zenodo	Scientific community, Industry	Worldwide	Worldwide
55	Other	Trowitzsch	NIGENS anechoic earsignals	12/2016	Zenodo	Scientific community	Worldwide	Worldwide

List of dissemination activities								
No	Type of activity[1]	Main leader	Title	Date/period	Place	Type of audience[2]	Size of audience	Countries addressed
56	Invited talk in special session	Blauert	Tool for instrumental assessment of the quality of experience original and simulated spaces for acoustic performances	June 2017, to come	Acoustics ‘17 Joint conf. EAA and ASA)	Acousticians in Virtual Reality	ca. 100 worldwide	Boston MA
57	Distinguished plenary lecture,	Blauert	Reading the world with two ears	July 2017 to come	24th Intl. Congr. Sound Vibr.	Acousticians in all fields	ca. 500 worldwide	United Kingdom, London

[1] A drop down list allows choosing the dissemination activity: publications, conferences, workshops, web, press releases, flyers, articles published in the popular press, videos, media briefings, presentations, exhibitions, thesis, interviews, films, TV clips, posters, Other.
[2] A drop down list allows choosing the type of public: Scientific Community (higher education, Research), Industry, Civil Society, Policy makers, Medias, Other ('multiple choices' is possible).

Table B2

Type of exploitable foreground *	Description of exploitable foreground	Confidential Y/N	Foreseen embargo date (dd/mm/yy)	Exploitable product(s) or measure(s)	Sector(s) of application **	Timetable, commercial or any other use	Patents or other IPR exploitation (licences)	Owner & other beneficiary(s) involved
General advancement of knowledge	Database of acoustic scenes and measurements	N	N/A	Database of acoustic scenes and associated measurements, to support development and testing of future audio technologies	M72.1 - Research and experimental development on natural sciences and engineering	Available from November 2016	None	Two!Ears consortium
General advancement of knowledge	Auditory front-end software (AFE)	N	N/A	Open source software for auditory modelling	M72.1 - Research and experimental development on natural sciences and engineering	Available from November 2016	None	Two!Ears consortium
General advancement of knowledge	Blackboard software, knowledge sources and associated software framework for machine learning	N	N/A	Open source software for acoustic scene analysis, incorporating acoustic processing and blackboard software architecture	M72.1 - Research and experimental development on natural sciences and engineering	Available from November 2016	None	Two!Ears consortium

Type of exploitable foreground *	Description of exploitable foreground	Confidential Y/N	Foreseen embargo date (dd/mm/yy)	Exploitable product(s) or measure(s)	Sector(s) of application **	Timetable, commercial or any other use	Patents or other IPR exploitation (licences)	Owner & other beneficiary(s) involved
General advancement of knowledge	Robotic unit for KEMAR head rotation (mechanical design and software)	N	N/A	Silent controllable device allowing neck rotation of the KEMAR manikin	M72.1 - Research and experimental development on natural sciences and engineering J59.2 - Sound recording and music publishing activities	Available from November 2016	None	Two!Ears consortium
General advancement of knowledge	Stereo-vision system for KEMAR or other dummy heads (hardware and software)	N	N/A	Hard and software for stereoscopic data acquisition and streaming to be mounted on KEMAR or similar Head-and-Torso-Simulators (HATS)	M72.1 - Research and experimental development on natural sciences and engineering J59.2 - Sound recording and music publishing activities	Available now	None	Two!Ears consortium

Type of exploitable foreground *	Description of exploitable foreground	Confidential Y/N	Foreseen embargo date (dd/mm/yy)	Exploitable product(s) or measure(s)	Sector(s) of application **	Timetable, commercial or any other use	Patents or other IPR exploitation (licences)	Owner & other beneficiary(s) involved
General advancement of knowledge	Robotics software and bridge	N	N/A	Binaural audio stream server (BASS) allowing efficient wireless streaming of audio from a binaural transducer	M72.1 - Research and experimental development on natural sciences and engineering J59.2 - Sound recording and music publishing activities	Available from November 2016	None	Two!Ears consortium
General advancement of knowledge	Low-cost Turtlebot solution for active exploration in an educational context	N	N/A	Software and hardware for audio recording and motor control of Turtlebot robot, allowing active exploration to be researched/demonstrated at low cost in educational institutions	M72.1 - Research and experimental development on natural sciences and engineering	Available now	None	Two!Ears consortium

Type of exploitable foreground *	Description of exploitable foreground	Confidential Y/N	Foreseen embargo date (dd/mm/yy)	Exploitable product(s) or measure(s)	Sector(s) of application **	Timetable, commercial or any other use	Patents or other IPR exploitation (licences)	Owner & other beneficiary(s) involved
General advancement of knowledge	Virtual test environment	N	N/A	A virtual test environment (LVTE) which simulates a robot and acoustic/video sensors in a 2D environment (based on BEFT)	M72.1 - Research and experimental development on natural sciences and engineering Application in robotic system for scene analysis and action planning	Available now	None	Two!Ears consortium
General advancement of knowledge	Test methods for quality evaluation of high-end systems	N	N/A	Specification of test methods for high-quality audio reproduction systems and statistical methods for data analysis, including freely available source data and processing information: - Localization - Coloration - Preference in pair-wise comparison tests	M72.1 - Research and experimental development on natural sciences and engineering J59.2 - Sound recording and music publishing activities	Available now	None	Two!Ears consortium

Type of exploitable foreground *	Description of exploitable foreground	Confidential Y/N	Foreseen embargo date (dd/mm/yy)	Exploitable product(s) or measure(s)	Sector(s) of application **	Timetable, commercial or any other use	Patents or other IPR exploitation (licences)	Owner & other beneficiary(s) involved
General advancement of knowledge	Software components for audio quality assessment	N	N/A	Software components as part of Two!Ears software that enable prediction of - Localization - Coloration - Pair-wise preference	M72.1 - Research and experimental development on natural sciences and engineering J59.2 - Sound recording and music publishing activities	Publicly available now; preference components available approximately mid January 2017	None	Two!Ears consortium

In addition to the table, please provide a text to explain the exploitable foreground, in particular:

- Its purpose
- How the foreground might be exploited, when and by whom
- IPR exploitable measures taken or intended
- Further research necessary, if any
- Potential/expected impact (quantify where possible)

General remark: In all cases of foreground, further research is foreseen to carry on the successful work of TWO!EARS. As a FET-Open project, its main goal was to foster science in auditory and audiovisual perception modelling, audio-visual signal processing, audio technology and robotics perception. All foreground items listed in the table and described in more detail below represent output that can be used in future research and technology development in the addressed areas.
For almost all of the foreground items, different application domains can be conceived, as described in more detail in Chapter 4 of the TWO!EARS Final Project Report. These applications include cocktail-party processors to perform efficient auditory-stream segregation in acoustically adverse conditions, auditory-object identification and tracking, automatic speaker identification and speech understanding, automatic perceptual analysis of auditory environments, perceptual audio processing comprising perceptual scene segregation, object extraction, perceptual coding and reproduction, models for Quality of Experience (QoE) for evaluating performance spaces, audio transmission and spatial reproduction, and robotics systems including human-like robot-audition.

Database of acoustic scenes and measurements The purpose of this database of acoustic scenes and associated measurements is to support development and testing of future audio technologies. License: CC BY-SA 3.0, see https://creativecommons.org/licenses/by-sa/3.0/. The database contains a number of listening test results as labels. Each entry is clearly described in terms of the underlying scenario and technical specification details. Thus, it can be exploited by any scientist or general person with interest in the field. Since the TWO!EARS project targets reproducible research, no measures beyond the provision and maintenance of the database are planned. By making it open access, the goal is that principally others may add their data as well so that further research becomes possible. The database forms a key part of the TWO!EARS mission. The scientific and technological impact can be relatively large since such databases are rare. Also, there is a thorough documentation available at http://docs.twoears.eu/en/latest/database/ so that usage is simplified.

Auditory front-end software (AFE) The purpose of this Open source software for auditory modelling is to enable a testbed functionality and as well as individual components for larger-scale machine learning approaches that shall use auditory features as input information. The AFE can support academic institutions as well as industry in testing and creating novel algorithms for modelling human auditory and audiovisual perception. The AFE uses MATLAB, which is used in industry and academia in audio and video signal processing contexts. The AFE software is published under GNU General Public License, version 3, see http://www.gnu.org/licenses/gpl-3.0.html. By making it open access via github (https://github.com/TWOEARS), the goal is that others can use it and also add their model components as well, so that further research becomes possible. The AFE forms a key part of the TWO!EARS mission. The scientific and technological impact can be relatively large since such open model frameworks that can operate on real audio data and not specific publication results are very rare. Also, there is a thorough documentation available at http://docs.twoears.eu/en/latest/afe/ so that usage is simplified. The model can be integrated with TWO!EARS Binaural Simulator (see Section 3.1 of the TWO!EARS Final Project Report and the documentation here: http://docs.twoears.eu/en/latest/binsim/).

Blackboard software, knowledge sources and associated software framework for machine learning. The purpose of this foreground is to provide a coherent software framework for audiovisual and cognitive modelling. It might be exploited by researchers in machine hearing, audio-visual processing, audio information extraction, robotics, music analysis and machine learning. It is sufficiently general to be applied to other research applications also. The software has been released as open source (https://github.com/TWOEARS). No further IPR exploitation is planned; this is a tool that we are making available to the research community in the interests of stimulating further scientific progress. We are aware that other labs outside of the Two!Ears consortium are already using the software, including Dr. Antoine Deleforge of the major French research institute INRIA.

Robotic unit for KEMAR head rotation *and* dedicated stereovision unit (mechanical design and software). The purpose of these two items of foreground is to provide a means of rotating the head of a KEMAR acoustic manikin, allowing research on head movements and sound collection in which the manikin head is automated on the one hand, and of capturing, streaming and processing additional binocular / stereoscopic visual data with the same HATS. An important aspect of the foreground on head rotation is that it can be fitted as a replacement collar to a KEMAR and the motor is absolutely silent. It might be exploited by researchers in machine hearing, robotics and acoustic analysis. The software and design are available as open source and no further IPR exploitation is planned. It uses ROS and GenoM3, hence is under BSD 2-Clause License. The impact will be on a number of academic and commercial users of KEMAR who wish to incorporate head movements and/or stereoscopic vision in their recordings.

Robotics software and bridge. The purpose of the foreground is to provide a cost-effective, efficient and reliable means of capturing binaural audio recordings from a mobile robot and streaming it to a computer for analysis. It will be useful to researchers in robotics, music recording, acoustics and machine hearing. The software and hardware required are available as open source or as low-cost widely-used components (e.g. Raspberry Pi). No further IPR exploitation is planned. The impact will be on academic, commercial and educational users who wish to make sound recordings from a mobile robot platform. The low cost, in particular, will make it accessible to educational users.

Low-cost Turtlebot solution for active exploration in an educational context. The purpose of this foreground is to provide a cost-effective platform for research and teaching in the field of active exploration. It combines a low-cost commercial robot (Turtlebot) with software that allows audio capture and motion control linked to the blackboard system described above. It will be particularly useful for educational establishments who wish to address cutting-edge topics in mobile robotics, while also being simple enough for use in schools and colleges. No IPR exploitation is planned; we have made the software and hardware designs freely available. The impact will primarily be on educational users in schools, colleges and universities.

Virtual test environment. The purpose of this foreground is to provide a virtual environment for designing and testing active exploration strategies in robotics. The simulator allows a virtual robot to be moved around a 2D environment, with sensor input such as audio and virtual cameras available. It might be exploited by researchers in robotics and machine hearing. The software has been released as open source. No further IPR exploitation is planned; this is a tool that we are making available to the research community in the interests of stimulating further scientific progress.

* A drop down list allows choosing the type of foreground: General advancement of knowledge, Commercial exploitation of R&D results, Exploitation of R&D results via standards, exploitation of results through EU policies, exploitation of results through (social) innovation.

** A drop down list allows choosing the type sector (NACE nomenclature) : http://ec.europa.eu/competition/mergers/cases/index/nace_all.html

Appendix II: Report on societal implications

Report on societal implications

Replies to the following questions will assist the Commission to obtain statistics and indicators on societal and socio-economic issues addressed by projects. The questions are arranged in a number of key themes. As well as producing certain statistics, the replies will also help identify those projects that have shown a real engagement with wider societal issues, and thereby identify interesting approaches to these issues and best practices. The replies for individual projects will not be made public.

A General Information *(completed automatically when **Grant Agreement number** is entered.*

Grant Agreement Number:	618075
Title of Project:	TWO!EARS
Name and Title of Coordinator:	Alexander Raake, Prof. Dr.-Ing.

B Ethics

1. Did your project undergo an Ethics Review (and/or Screening)?	x
• If Yes: have you described the progress of compliance with the relevant Ethics Review/Screening Requirements in the frame of the periodic/final project reports?	x
Special Reminder: the progress of compliance with the Ethics Review/Screening Requirements should be described in the Period/Final Project Reports under the Section 3.2.2 *'Work Progress and Achievements'*	Note: This has been reported under Section 3.3.3.1 Ethics considerations
2. Please indicate whether your project involved any of the following issues (tick box) :	x
RESEARCH ON HUMANS	
• Did the project involve children?	0
• Did the project involve patients?	0
• Did the project involve persons not able to give consent?	0
• Did the project involve adult healthy volunteers?	x
• Did the project involve Human genetic material?	0
• Did the project involve Human biological samples?	0
• Did the project involve Human data collection?	x
RESEARCH ON HUMAN EMBRYO/FOETUS	
• Did the project involve Human Embryos?	0
• Did the project involve Human Foetal Tissue / Cells?	0
• Did the project involve Human Embryonic Stem Cells (hESCs)?	0
• Did the project on human Embryonic Stem Cells involve cells in culture?	0
• Did the project on human Embryonic Stem Cells involve the derivation of cells from Embryos?	0
PRIVACY	
• Did the project involve processing of genetic information or personal data (eg. health, sexual lifestyle, ethnicity, political opinion, religious or philosophical conviction)?	0

• Did the project involve tracking the location or observation of people?	x[1]
RESEARCH ON ANIMALS	
• Did the project involve research on animals?	0
• Were those animals transgenic small laboratory animals?	0
• Were those animals transgenic farm animals?	0
• Were those animals cloned farm animals?	0
• Were those animals non-human primates?	0
RESEARCH INVOLVING DEVELOPING COUNTRIES	
• Did the project involve the use of local resources (genetic, animal, plant etc)?	0
• Was the project of benefit to local community (capacity building, access to healthcare, education etc)?	0
DUAL USE	
• Research having direct military use	0
• Research having the potential for terrorist abuse	0

C Workforce Statistics

3. Workforce statistics for the project: Please indicate in the table below the number of people who worked on the project (on a headcount basis).

Type of Position	Number of Women	Number of Men
Scientific Coordinator	0	1
Work package leaders	0	8
Experienced researchers (i.e. PhD holders)	1	20
PhD Students	0	5
Other	2	

4. How many additional researchers (in companies and universities) were recruited specifically for this project?	7
Of which, indicate the number of men:	7

[1] Location of test participant's head tracked in some of the tests conducted under multimedia test laboratory settings.

D Gender Aspects

5. Did you carry out specific Gender Equality Actions under the project?	○ X	Yes No

6. Which of the following actions did you carry out and how effective were they?

		Not at all effective				Very effective
❑	Design and implement an equal opportunity policy	○	○	○	○	○
❑	Set targets to achieve a gender balance in the workforce	○	○	○	○	○
❑	Organise conferences and workshops on gender	○	○	○	○	○
❑	Actions to improve work-life balance	○	○	○	○	○

○ Other: n/a

7. Was there a gender dimension associated with the research content – i.e. wherever people were the focus of the research as, for example, consumers, users, patients or in trials, was the issue of gender considered and addressed?

X Yes- please specify: It was attempted that approximately 50% of test subjects would female, 50% male to address a good average sample of the target population.

○ No

E Synergies with Science Education

8. Did your project involve working with students and/or school pupils (e.g. open days, participation in science festivals and events, prizes/competitions or joint projects)?

X Yes- please specify: University students and pupils as visitors for : Long Night of Science Berlin 2014, Nuit Européenne des Chercheurs Toulouse, 2016
University students (PhD candidates) Two !Ears Summer School 2015, Toulouse

○ No

9. Did the project generate any science education material (e.g. kits, websites, explanatory booklets, DVDs)?

X Yes- please specify: Website, Video, documented software

○ No

F Interdisciplinarity

10. Which disciplines (see list below) are involved in your project?

X Main discipline[2]: 1.1, 2.2, 5.1

X Associated discipline[2]: 3.3 | ○ Associated discipline[2]:

G Engaging with Civil society and policy makers

11a Did your project engage with societal actors beyond the research community? *(if 'No', go to Question 14)*	X ○	Yes No

11b If yes, did you engage with citizens (citizens' panels / juries) or organised civil society (NGOs, patients' groups etc.)?

○ No
○ Yes- in determining what research should be performed
○ Yes - in implementing the research
X Yes, in communicating /disseminating / using the results of the projecte

[2] Insert number from list below (Frascati Manual).

11c	**In doing so, did your project involve actors whose role is mainly to organise the dialogue with citizens and organised civil society (e.g. professional mediator; communication company, science museums)?**	O X	Yes No

12. Did you engage with government / public bodies or policy makers (including international organisations)

- O No
- O Yes- in framing the research agenda
- O Yes - in implementing the research agenda
- X Yes, in communicating /disseminating / using the results of the project

13a Will the project generate outputs (expertise or scientific advice) which could be used by policy makers?

- X Yes – as a **primary** objective (please indicate areas below- multiple answers possible) → "P"
- X Yes – as a **secondary** objective (please indicate areas below - multiple answer possible) → "S"
- O No

13b If Yes, in which fields?

Agriculture Audiovisual and Media **-- S** Budget Competition Consumers Culture Customs Development Economic and Monetary Affairs Education, Training, Youth Employment and Social Affairs		Energy Enlargement Enterprise Environment External Relations External Trade Fisheries and Maritime Affairs Food Safety Foreign and Security Policy Fraud Humanitarian aid		Human rights Information Society **-- P** Institutional affairs Internal Market Justice, freedom and security Public Health **-- P** Regional Policy Research and Innovation **-- P** Space Taxation Transport	

13c If Yes, at which level? ○ Local / regional levels ○ National level ○ European level X International level	
H Use and dissemination	
14. How many Articles were published/ accepted for publication in peer-reviewed journals?	**9**
To how many of these is open access[3] provided?	**7**
How many of these are published in open access journals?	**6**
How many of these are published in open repositories?	**1**
To how many of these is open access not provided?	**2**
Please check all applicable reasons for not providing open access:	
❑ publisher's licensing agreement would not permit publishing in a repository ❑ no suitable repository available ❑ no suitable open access journal available ❑ no funds available to publish in an open access journal X lack of time and resources ❑ lack of information on open access ❑ other[4]: ……………	

15. How many new patent applications ('priority filings') have been made? *("Technologically unique": multiple applications for the same invention in different jurisdictions should be counted as just one application of grant).*		n/a
16. Indicate how many of the following Intellectual Property Rights were applied for (give number in each box).	Trademark	n/a
	Registered design	n/a
	Other	n/a
17. How many spin-off companies were created / are planned as a direct result of the project?		n/a
Indicate the approximate number of additional jobs in these companies:		

18. Please indicate whether your project has a potential impact on employment, in comparison with the situation before your project:

X	Increase in employment, or	X	In small & medium-sized enterprises
X	Safeguard employment, or	X	In large companies
❑	Decrease in employment,	❑	None of the above / not relevant to the project
❑	Difficult to estimate / not possible to quantify		

19. For your project partnership please estimate the employment effect resulting directly from your participation in Full Time Equivalent (*FTE = one person working fulltime for a year*) jobs: Difficult to estimate / not possible to quantify → only academic partners	*Indicate figure:* 0.5 – 1 per partner based on follow-up funding. With 9 partners : 4.5-9

[3] Open Access is defined as free of charge access for anyone via Internet.
[4] For instance: classification for security project.

I Media and Communication to the general public

20. As part of the project, were any of the beneficiaries professionals in communication or media relations?

○ Yes X No

21. As part of the project, have any beneficiaries received professional media / communication training / advice to improve communication with the general public?

X Yes ○ No

22 Which of the following have been used to communicate information about your project to the general public, or have resulted from your project?

X	Press Release	X	Coverage in specialist press
❑	Media briefing	X	Coverage in general (non-specialist) press
X	TV coverage / report	X	Coverage in national press
X	Radio coverage / report	❑	Coverage in international press
X	Brochures /posters / flyers	X	Website for the general public / internet
X	DVD /Film /Multimedia	X	Event targeting general public (festival, conference, exhibition, science café)

23 In which languages are the information products for the general public produced?

X	Language of the coordinator: German	X	English
X	Other language(s): French		

Question F-10: Classification of Scientific Disciplines according to the Frascati Manual 2002 (Proposed Standard Practice for Surveys on Research and Experimental Development, OECD 2002):

FIELDS OF SCIENCE AND TECHNOLOGY

1. NATURAL SCIENCES

1.1 Mathematics and computer sciences [mathematics and other allied fields: computer sciences and other allied subjects (software development only; hardware development should be classified in the engineering fields)]

1.2 Physical sciences (astronomy and space sciences, physics and other allied subjects)

1.3 Chemical sciences (chemistry, other allied subjects)

1.4 Earth and related environmental sciences (geology, geophysics, mineralogy, physical geography and other geosciences, meteorology and other atmospheric sciences including climatic research, oceanography, vulcanology, palaeoecology, other allied sciences)

1.5 Biological sciences (biology, botany, bacteriology, microbiology, zoology, entomology, genetics, biochemistry, biophysics, other allied sciences, excluding clinical and veterinary sciences)

2 ENGINEERING AND TECHNOLOGY

2.1 Civil engineering (architecture engineering, building science and engineering, construction engineering, municipal and structural engineering and other allied subjects)

2.2 Electrical engineering, electronics [electrical engineering, electronics, communication engineering and systems, computer engineering (hardware only) and other allied subjects]

2.3. Other engineering sciences (such as chemical, aeronautical and space, mechanical, metallurgical and materials engineering, and their specialised subdivisions; forest products; applied sciences such as geodesy, industrial chemistry, etc.; the science and technology of food production; specialised technologies of interdisciplinary fields, e.g. systems analysis, metallurgy, mining, textile technology and other applied subjects)

3. MEDICAL SCIENCES

3.1 Basic medicine (anatomy, cytology, physiology, genetics, pharmacy, pharmacology, toxicology, immunology and immunohaematology, clinical chemistry, clinical microbiology, pathology)

3.2 Clinical medicine (anaesthesiology, paediatrics, obstetrics and gynaecology, internal medicine, surgery, dentistry, neurology, psychiatry, radiology, therapeutics, otorhinolaryngology, ophthalmology)

3.3 Health sciences (public health services, social medicine, hygiene, nursing, epidemiology)

4. AGRICULTURAL SCIENCES

4.1 Agriculture, forestry, fisheries and allied sciences (agronomy, animal husbandry, fisheries, forestry, horticulture, other allied subjects)

4.2 Veterinary medicine

5. SOCIAL SCIENCES

5.1 Psychology

5.2 Economics

5.3 Educational sciences (education and training and other allied subjects)

5.4 Other social sciences [anthropology (social and cultural) and ethnology, demography, geography (human, economic and social), town and country planning, management, law, linguistics, political sciences, sociology, organisation and methods, miscellaneous social sciences and interdisciplinary , methodological and historical S1T activities relating to subjects in this group. Physical anthropology, physical geography and psychophysiology should normally be classified with the natural sciences].

6. HUMANITIES

6.1 History (history, prehistory and history, together with auxiliary historical disciplines such as archaeology, numismatics, palaeography, genealogy, etc.)

6.2 Languages and literature (ancient and modern)

6.3 Other humanities [philosophy (including the history of science and technology) arts, history of art, art criticism, painting, sculpture, musicology, dramatic art excluding artistic "research" of any kind, religion, theology, other fields and subjects pertaining to the humanities, methodological, historical and other S1T activities relating to the subjects in this group]