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

Deliverable 1.3

Final database of audio-visual scenarios

WP 1 *

November 30, 2016

* 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 package:	1
Document number:	D1.3
Document title:	Final database of audio-visual scenarios
Version:	1.0

Delivery date:	30. November 2016
Actual publication date:	01. December 2016
Dissemination level:	Public
Nature:	Other

Editor(s):	Sascha Spors
Author(s):	Fiete Winter, Hagen Wierstorf, Ivo Trowitzsch
Reviewer(s):	Jonas Braasch, Dorothea Kolossa, Bruno Gas, Klaus Obermayer

About this Document

This document is split in two parts. Part I contains the report on the final database including the achievements made during the whole TWO!EARS project. Part II is the Portable Document Format (PDF) version of the online documentation of the TWO!EARS public database[1] including all datasets made available to the community. Generating either webpages or PDF documents is possible due to the markup language named reStructuredText (reST) and the corresponding interpreter sphinx[2] used for the TWO!EARS documentation. The first and the second part can be recognized by their respective arabic and capital roman page numbering.

1 online version available at `http://docs.twoears.eu/en/latest/`

2 see `http://www.sphinx-doc.org`

Part I

Report on Final Database

Contents of Report (Part I)

1 Executive Summary **7**

2 Simulation Framework **9**
2.1 Techniques for Binaural Synthesis 9
2.1.1 Pre-Recorded Binaural Signals 9
2.1.2 Static Head-Related and Binaural Room Impulse Responses 10
2.1.3 Dynamic Binaural Room Impulse Responses 11
2.1.4 Data-Based Binaural Synthesis 12
2.1.5 Numerical Simulation of Acoustic Environments 13
2.1.6 Comparison of Binaural Synthesis Techniques 14
2.2 Synthesis of Ear Signals 15
2.2.1 SoundScape Renderer 15
2.2.2 Acoustic Scene 16
2.2.3 Configuration and Interfaces 17
2.2.4 Integration and Application 18
2.3 Simulation of Visual Stimuli 18

3 Accessibility and Compatibility of the Database **19**
3.1 Infrastructure 19
3.1.1 Public and Project-Internal Databases 19
3.1.2 Software Interface for Public Database 20
3.2 Data Formats 22
3.2.1 Impulse Responses 22
3.2.2 Audio-Visual Scene Description 22

4 Conclusions and Outlook **25**

Bibliography **27**

1 Executive Summary

The acoustic signals at the ears serve as input for the auditory scene analysis performed by the human auditory system. The same holds for the human visual system where the eyes provide the input. The goal of the Two!Ears project is to develop an intelligent, 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 and eye images 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. For the synthesis a decent amount of recorded and measured data has to be provided. Furthermore, perceptual labels are mandatory, as the computational model has to be evaluated against human performance. This calls for a central database in order to provide access to this data among the members of the consortium and the public.

Within Work Package 1, an audio-visual simulation framework was developed. An overview on the available acoustic simulations techniques and technical details about the implementation are given in Chapter 2.

In the Deliverable D1.1, a hybrid infrastructure separating the publishable, open source licensed content from the restricted, project internal data has been reported. During the second year of the Two!Ears project the infrastructure of the project internal database was adjusted to encounter problems related to the significant growth of the database. Finally additional functionalities for a more straightforward access to the data have been added during the third year of the project. The final infrastructure and interfaces are documented in Chapter 3.

2 Simulation Framework

This section provides a brief overview on the methods applied in Two!Ears for synthesizing ear-signals. Each technique has particular benefits and weaknesses which are summarized in Section 2.1.6. Depending on the task, one particular technique or a combination is used for the synthesis of ear signals. Section 2.2 reports on the technical details of the audio simulation framework. The simulation of visual stimuli is reviewed in Section 2.3 together with the progress in the processing of visual stimuli.

2.1 Techniques for Binaural Synthesis

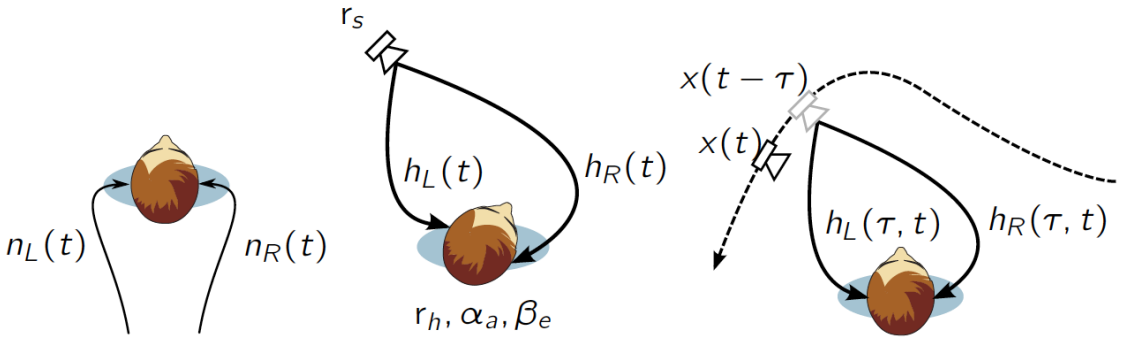
Binaural synthesis refers to the synthesis of the sound pressure at a defined position in the ear-canal. Often, the sound pressure at the blocked entrance of the ear canal is used as reference. This is also the reference assumed in Two!Ears.

2.1.1 Pre-Recorded Binaural Signals

A straightforward approach is to record the ear signals. The recording can either be performed by placing small microphones at a defined position in the ear canal of a listener. Intuitively the microphone should be placed close to the eardrum. However due to the medical risks involved in such a procedure, binaural signals are often captured at the blocked entrance of the ear canal. This has proven to work well in practice. As an alternative to a human listener, such recordings are also be performed by a Head and Torso Simulator (HATS).

The synthesis of pre-recorded binaural signals is performed by playing back the recorded signals. This is illustrated in Figure 2.1a. A (free-field or diffuse-field) compensation of the in-ear microphones or HATS frequency response might be required to compensate for their influences.

Any sound field can be captured. This also includes diffuse sound fields and moving sound sources. However, the head orientation is fixed by the recording and cannot be changed during synthesis. The Two!Ears model is a binaural model which includes feedback



(a) pre-recorded binaural signals **(b)** static binaural room impulse responses **(c)** dynamic binaural room impulse responses

Figure 2.1: Techniques for binaural synthesis applied in TWO!EARS. Left/right binaural signals are denoted by $n_{\{L,R\}}(t)$, BRIRs by $h_{\{L,R\}}(t)$, time-variant BRIRs by $h_{\{L,R\}}(\tau, t)$. The source signal is denoted by $x(t)$, the source and receiver position by r_s and r_h, respectively.

mechanisms and active listening. Therefore, the synthesis of pre-recorded binaural signals is only of interest for diffuse background noise where the head orientation does not play a role. Table 2.1 summarizes the features of binaural synthesis using pre-recorded binaural signals.

WP 1 has reviewed existing databases of pre-recorded binaural signals and collected them in the central database. See D 1.1 for a listing of the included material. The simulation framework for the computation of ear signals allows the use of pre-recorded binaural signals. For a detailed description refer to Section 2.2.

2.1.2 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). HRIRs depend on the source position with respect to the listener and on the head orientation. BRIRs depend additionally on the position and orientation of the source and the listener in the environment. In order to capture these degrees of freedom, databases of HRIR/BRIRs have to be captured. For HRIRs typically the incidence angle of the source for a fixed distance is varied. For BRIRs the head orientation is varied for a fixed source and listener position.

The synthesis of ear-signals is performed by convolving the desired source signal with the appropriate left/right HRIR/BRIR from the database. This is illustrated in Fig-

ure 2.1b. A change in head-orientation can be considered by exchanging the set of HRIR/BRIRs.

The use of HRIR/BRIRs is restricted to compact sound sources which do not move. Diffuse sound fields can be approximated by superposition of many sources from different directions. Moving sound sources can be approximated by concatenating a series of static source positions. However, this is typically not possible for BRIRs since this would require a densely captured grid of source positions. A change in head-orientation can be modeled by exchanging the HRIR/BRIRs. Hence, static HRIR/BRIRs are used for the simulation of single or some few sources. The properties of this technique are summarized in Table 2.1.

WP 1 has reviewed existing databases of HRIR/BRIRs, converted and collected them in the central database. See Part II for a listing of the included material. The simulation framework for the computation of ear signals allows the use of HRIR/BRIRs for the simulation of sound sources. For a detailed description refer to Section 2.2.

2.1.3 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. Alternatively, the time-variance of an environment can be captured for a fixed source and head position/orientation. Different head-orientations could be captured if the source trajectory were exactly reproducible. However that is hard to realize in practice.

The synthesis of ear-signals is performed by time-variant convolution of the desired source signal with the time-variant BRIRs, as illustrated in Figure 2.1c. The speed with which the sound source moves along the trajectory can be modified by subsampling or interpolation of the time-variant BRIRs.

The time-variant BRIRs contain the fine structure of the room which is exited by the moving sound source. All effects due to the moving source, e.g. the Doppler effect, are included. In practice, only a fixed head-orientation and trajectory can be considered. This technique is suitable for the accurate synthesis of moving sound sources in rooms. Its properties are summarized in Table 2.1.

In the past, time variant system identification techniques have been applied to the measurement of spatially continuous HRIRs [1]. WP 1 has extended this by considering moving sound sources and time-variant acoustic environments. This required an advancement

of the existing techniques [2, 3]. Various simulations and measurements have been performed to evaluate the technique in synthetic and real environments. Figure 2.2 shows time-variant BRIRs for a moving scatterer which obstructs the direct path between the loudspeaker and HATS used for the measurement. This constitutes a time-variant acoustic scenario. The effects caused by the scattering are clearly visible in the BRIRs. Future work includes the inclusion of the method into the simulation framework described in Section 2.2.

2.1.4 Data-Based Binaural Synthesis

The binaural synthesis techniques above cannot cope with translatory movements of the listener in reverberant environments. Small scale translatory movements are considered to provide additional localization cues. 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 [4, 5]. 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

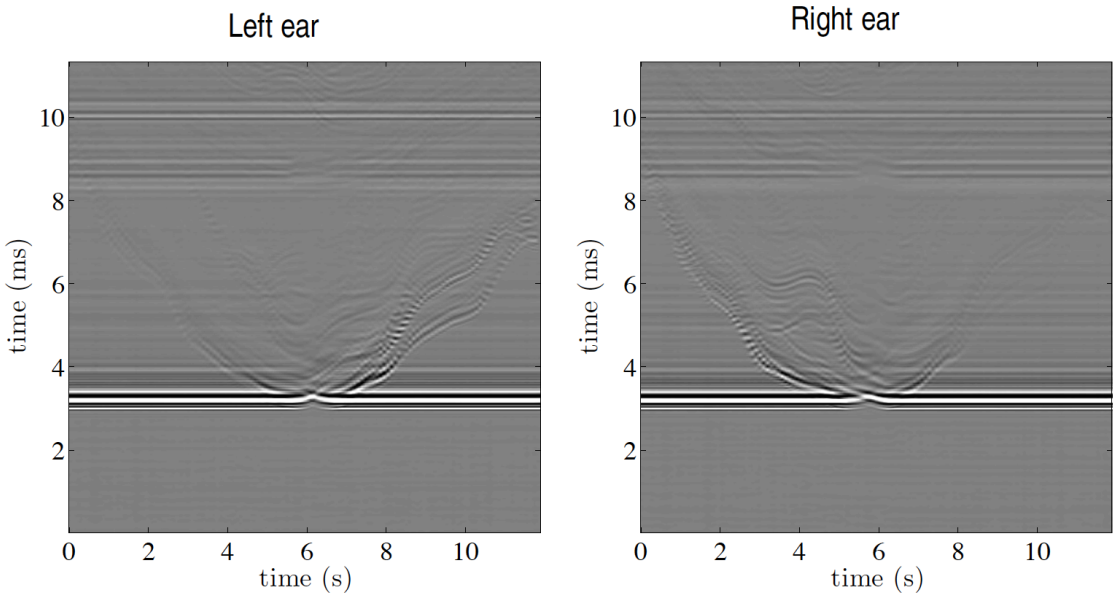


Figure 2.2: Dynamic BRIRs captured in a typical office room using a static loudspeaker and HATS position. The horizontal axis denotes the time t a BRIR has been captured. The vertical axis is the time axis τ of the BRIR. A vertical slice represents a BRIR at a given time. A moving scatterer is obstructing the direct path between loudspeaker and HATS at $t \approx 6$ s.

up for auralization. The overall procedure is illustrated in Figure 2.3. The technique allows the use of individual HRIRs. As an alternative to the procedure outlined above, Multi-Channel Room Impulse Responses (MRIRs) from a source to the microphone array can be used as input. 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 accuracy depends, amongst others, on the number of microphones used, their self-noise and the translation distance. The properties of the technique are reviewed in Table 2.1.

The basic processing for data-based binaural synthesis has been published by URO. The perceptual properties in terms of localization for various technical parameters have been evaluated in WP 1 and published [6].

2.1.5 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. State-of-the art simulation software combines

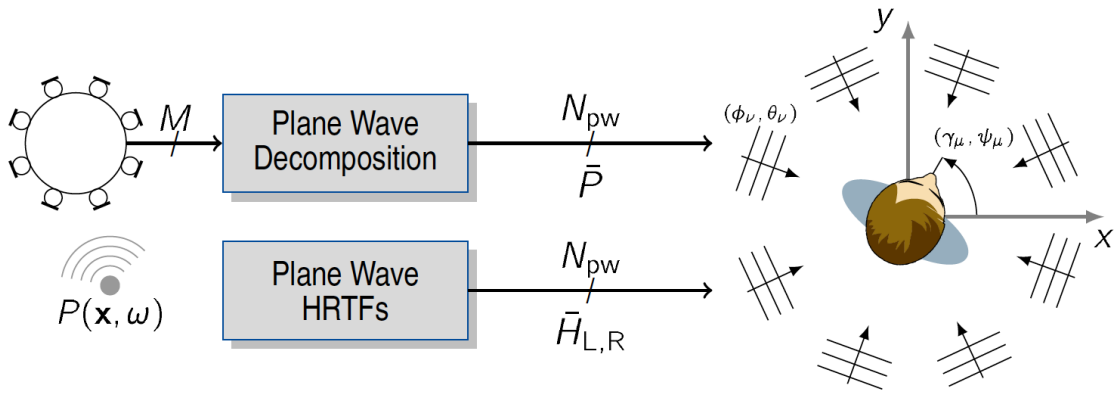


Figure 2.3: Data-based binaural synthesis including small scale translatory movements of the listener. The number of microphones is denoted by M, the number of plane waves by N_{pw}, the coefficients of the plane wave decomposition by $\bar{P}$, the left/right far-field HRTFs by $\bar{H}_{\text{L,R}}$. The incidence angles of the plane waves are denoted by (ϕ_ν, θ_ν), the look directions of the listener by (γ_μ, ψ_μ).

a mirror image source model for the direct sound and early-reflections with energetic methods (e.g. ray tracing) for the late reverberation. Hence, diffraction and standing waves are often neglected or only approximated to some limited degree. Another source of inaccuracies is the handling and determination of boundary conditions. For an accurate numerical simulation, the incidence- and frequency-dependent reflection properties of all acoustically relevant materials for a given environment are required. In practice these are only known to a limited degree. Most of the available software packages allow to compute BRIRs for a given environment.

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. However, the physical accuracy of such simulations is limited due to the underlying approximations and material properties. It is not clear how relevant acoustic cues, e.g. for localization, are preserved. The properties of the technique are listed in Table 2.1.

WP 1 has screened most of the available commercial and non-commercial simulation frameworks for the application in TWO!EARS. Besides accuracy, the software-based control of source and listener position/orientation is a major selection criterion, due to the intended exploratory nature of the TWO!EARS model. The simulation framework RAVEN of the Institute of Technical Acoustics, RWTH Aachen [7] features a MATLAB interface to control these and other parameters. This allows a seamless integration into the ear-signal simulation framework. This track was not followed in TWO!EARS.

2.1.6 Comparison of Binaural Synthesis Techniques

Table 2.1 summarizes the properties of the different binaural synthesis techniques discussed above and their application in the TWO!EARS simulation framework.

technique	diffuse sound field	moving sources	head-orientation	head-translation	realism	application
pre-recorded signals (Sec. 2.1.1)	yes	yes	**no**	**no**	high	background noise
static HRIR/BRIRs (Sec. 2.1.2)	limited	limited	yes	limited	high	static sound sources
dynamic BRIRs (Sec. 2.1.3)	**no**	yes	**no**	**no**	high	moving sound sources
data-based synthesis (Sec. 2.1.4)	yes	**no**	yes	small-scale	high	static sound sources
numerical simulation (Sec. 2.1.5)	limited	yes	yes	yes	limited	acoustic navigation

Table 2.1: Comparison of binaural synthesis techniques used in Two!Ears.

2.2 Synthesis of Ear Signals

A binaural simulation framework has been implemented in order to appropriately address the defined scenarios and provide a feasible testbed for an efficient development of the auditory model. It provides synthesized ear signals of virtual acoustic environments and supports active explorative feedback, as it is defined in the model architecture. For model training and validation purposes various instances of the same scenario can be realized straightforwardly.
The software architecture of the binaural simulator is depicted in Figure 2.4. The open source software SoundScape Renderer (SSR) [8] is used as the core of the simulation tool (red box). While the core is written in C++, the rest of the framework is implemented in MATLAB in order to ensure compatibility to the software of the auditory model. A detailed explanation of individual system components is given in the following sections.

2.2.1 SoundScape Renderer

The SSR is a tool for realtime spatial audio reproduction, supporting the (dynamic) binaural synthesis of spatial audio scenes. For free-field scenarios it utilizes a set of a-priori measured HRIRs to recreate the acoustic signals at the ears of the listener. The sound reproduction is implemented by filtering a dry source signal with the appropriate HRIR for the left and the right ear. These HRIRs are selected by the position of the sound source relative to the listener's head. Different source distances are obtained by adapting the source's sound volume and delay. Dynamic scenes and listener movements can be adressed due to the frame-based signal processing framework [9] of the SSR. This allows for cross-fading the HRIRs in each time frame in order to simulate dynamic scenes or head

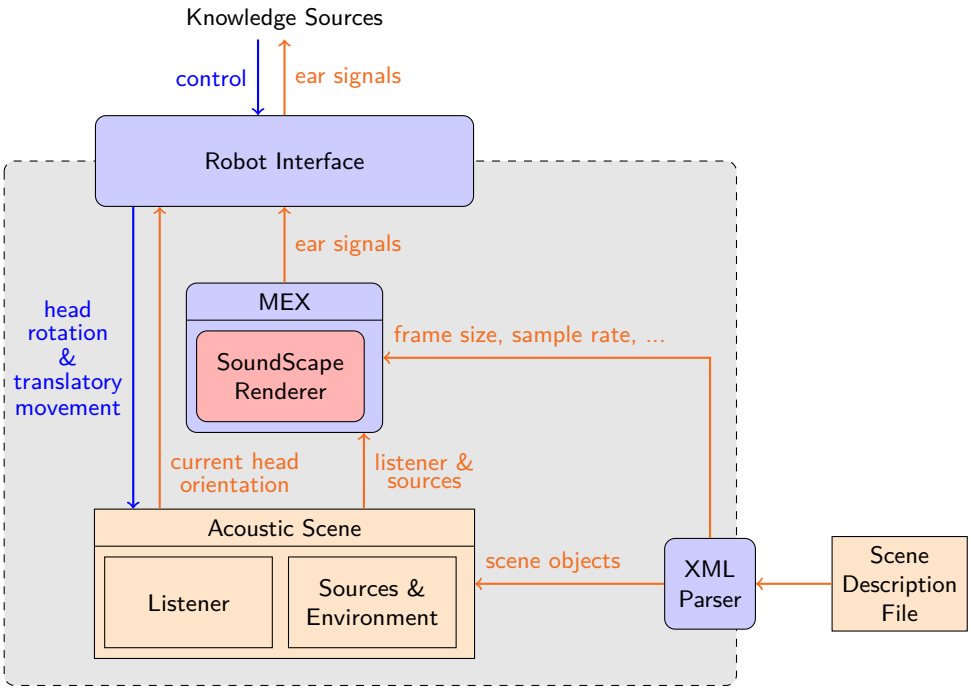


Figure 2.4: software architecture of the binaural simulator

motion.

The binaural synthesis of reverberant environments is handled using an individual set of BRIRs for each sound source. Since the respective source position and the room acoustics are already encoded in the impulse responses, scene dynamics are limited to the listener's head rotation. In addition, this rendering technique can be used to binaurally synthesize the sound reproduction of loudspeaker systems by taking the BRIRs of each loudspeaker into account.
The SSR can be controlled and configured via a MATLAB executable (MEX) interface (red and blue box, see fig. 2.4). While processing parameters like e.g. frame size and sample rate have to be defined initially and kept constant during a simulation run, properties of the listener's head and sound sources may be changed between each time frame.

2.2.2 Acoustic Scene

The acoustic scene (orange box, see fig. 2.4) is an ensemble of acoustic objects defined in a three-dimensional coordinate system. As a central object, the *Listener* represents the human head and acts as an acoustic sensor with two channels for the left and right

ear.

Sound *Sources* are essentially characterized by their position, source signal and source type. Several binaural reproduction techniques (see sec. 2.1) are selectable depending on the source type: Sources emitting a spherical or plane wave field in a non-reverberant environment are represented by *point* and *plane wave* sources, respectively. Binaural recordings of environments can be auralized via a two-channel *direct* source, while the plane wave decomposition (PWD) of multi-channel microphone array recordings are addressed by a multi-channel PWD source. The latter is implemented as group of plane wave sources, where each source is oriented and weighted according to the plane wave decomposition. The PWD source also allows for the simulation of spatially expanded noise sources.

Reverberant environments can be realized in two ways: As already shortly described in Section 2.2.1, each source may be endowed with a set of BRIRs. In this case source properties like e.g. position and source type are not configurable, since all physical information is already provided by the BRIRs. Active exploration by translatory movement of the listener is only possible, if the BRIR dataset includes different listening positions. As the second alternative, a image source model for rectangular rooms [10] is implemented within the binaural simulator. It simulates the specular reflections of sound at each wall and takes multiple reflections up to a user-defined order into account. For this purpose a *Room* object has to be defined.

2.2.3 Configuration and Interfaces

For the general application case the simulation core and the acoustic scene (see sections 2.2.1 and 2.2.2, respectively) are encapsulated (grey box, see Figure 2.4) and can only be manipulated and controlled via specific interfaces. In the following, these three methods are presented:

Scene Description File: The XML parsing functionality of the binaural simulation tool allows for the definition of the whole acoustic scene via an XML scene description file. This also includes events which manipulate a scene object at a certain point in time and describe the scene dynamics. The XML parser is based on the built-in W3C Document Object Model (DOM) of MATLAB. Besides the acoustic scene, also processing parameters of the simulation core can be initialized via the description file.

Robot Interface: In order to encapsulate the simulation functionalities and its specifics to the upper auditory model layers, a simple interface has been designed which narrows the possibilities for manipulation down to the control mechanisms of a real robot (blue arrows, see Figure 2.4). This strategy ensures compatibility between

the robotics platform and the simulation. Binaural signals of a desired length can be requested by the upper model stages. Explorative feedback is realized by head rotation and translatory movement, which implicitly modifies the head object of the acoustic scene.

Direct Configuration: For training and validation purposes of the auditory model, it is necessary to iterate over several parameter combinations and straightforwardly generate a large amount of training/validation scenarios. The definition of each scenario using the scene description file is not feasible for training and validation. The *expert* user therefore has the possibility to circumvent the encapsulation and directly access the acoustic scene and control the simulation core.

2.2.4 Integration and Application

The simulation framework for the synthesis of ear signals described above has been fully integrated into the TWO!EARS software framework. It is interfaced with the peripheral processing and feature extraction stage of work package two (WP2). The blackboard system, developed by WP3, acts as the core of the integrated framework and provides additional interfaces to include feedback mechanisms investigated within work package four (WP4) and a robotic interface to communicate with work package five (WP5). 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`.

The simulation framework has been applied successfully in the project. The full integration into all Workpackages has allowed significant scientific contributions in many tasks.

2.3 Simulation of Visual Stimuli

The Bochum Experimental Feedback Testbed (BEFT) allows to emulate scenarios of moderate complexity in a baseline VR environment. Synthetic footage can be acquired by using a virtual camera that is attached to the head of the emulated robot. For more details on visual processing in BEFT, refer to D 4.3, Section 5.

3 Accessibility and Compatibility of the Database

The synthesis of ear signals requires data characterizing the acoustic environment to be simulated. Physical and perceptual labels are required to systematically evaluate the model performance. This data is collected in a central database for seamless access by the model. The infrastructure of the database and data formats are discussed in the following.

3.1 Infrastructure

3.1.1 Public and Project-Internal Databases

The Two!Ears database represents a compilation of existing data from other databases and data recorded/measured in the project. Several datasets are published under an open source license allowing their re-distribution in the context of Two!Ears. However, for the development, testing and validation of the model additional data is necessary whose access is restricted to consortium members. A hybrid approach utilizing two databases, namely a project-internal and a public database, is therefore used. Both databases are endowed with a version control system allowing for the recovery of older file versions. In the first year the version control software git[1] has been favoured by the consortium. Due to the noticeable growth of both databases, an increasing number of consortium members reported problems as they tried to access the data via git. These problems were mostly related to the fact, that git is mainly designed to handle rather small text files, i.e. source code. Furthermore, the users had to download the entire database although they might only be interested in one file, as a so-called partial checkout is not straightforward in git. However, the version control mechanisms of git appear to be a very useful tool for the database administration.

All the needed data is stored in a restrictive, project-internal database, which relies on an

1 see `http://git-scm.com`

extension to git called git-media[1]. As shown in Fig. 3.1 the version control and storage of the data have been split up into a git repository for the meta data and a Secure Copy (SCP) file server. A hash id is computed for every binary file, which is added to the database. This id is stored in a text file located on the git repository having the same filename as the binary file. The binary file itself is uploaded to the file server. Whenever a binary file has been changed in the local copy of the user, the underlying algorithm recognizes this by comparing the hash id of the respective file with the hash id in the git repository. Furthermore, changes from other users are identified by changed ids in the git repository. Respective actions are then taken by the algorithm, i.e. download or upload a newer version of the binary file. In addition, the git repository contains the directory structure of the file server. The user therefore only needs to clone the complete git repository and then may decide which binary files or subdirectories should be downloaded from the file server.

The public database mirrors the open source licensed part of the data. It can be accessed via a web interface or with the version control software Apache Subversion (svn)[2]. As it has been decided amongst the project partners to make the public database available to the scientific community at an early stage of the project, the public database has been available after the first year of the Two!Ears project. It is currently hosted at TU Berlin[3] and its documentation is provided as part of the Two!Ears documentation[4]. A pdf version of this documentation is appended as part II of this document.

3.1.2 Software Interface for Public Database

For the dissemination among the scientific community and efficient work with the Two!Ears model, it is necessary to make the access to the public database as seamless as possible. While the version control support of svn is a huge benefit, it is primarily designed for cooperative software development mostly storing text and source code files on a server. Since the user needs all source code files in order to get the software working, svn provides a straightforward interface to download the whole repository with one command. For the mentioned application case the file sizes are small and download time is acceptable.
The Two!Ears database mostly contains binary files of comparably big size. Even for model applications where only a few files are needed, the whole database would have to be downloaded. This might distract potential users from testing/using the Two!Ears software framework. In order to overcome these limitations, a MATLAB software interface has been designed to access the needed files on demand. It circumvents the necessity of svn

1 see `https://github.com/alebedev/git-media`
2 see `https://subversion.apache.org/`
3 see `https://dev.qu.tu-berlin.de/projects/twoears-getdata/repository`
4 see `http://docs.twoears.eu/en/latest/`

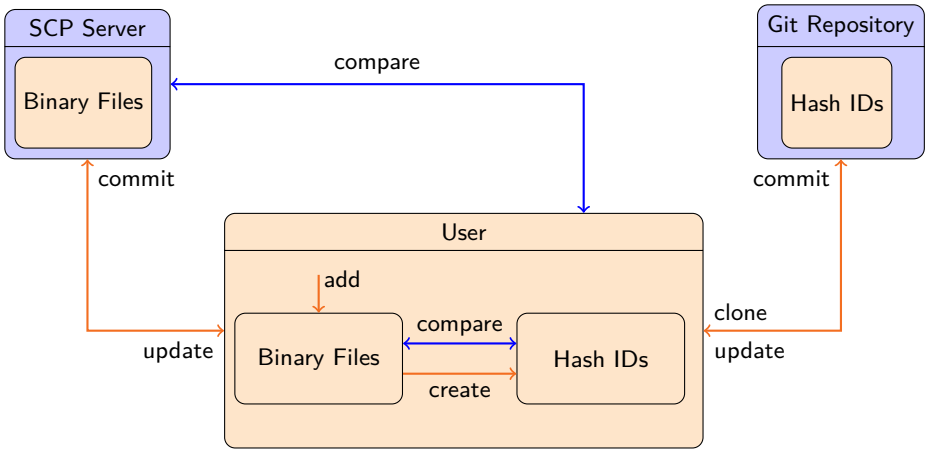


Figure 3.1: Structure of project-internal database

for downloading the files and uses the web interface of the database. Standard MATLAB functions[1] are used to access the remote data via the Hypertext Transfer Protocol (HTTP). The software interface provides a flexible file inclusion mechanism in order to prevent unnecessary downloads for the user. It takes different possible locations (in order of appearance) of the needed file into account:

1. Search the file relatively to the current working directory.
2. Search the file relatively to the root of the file system.
3. Search inside a local copy of the whole repository, which has been downloaded via svn. The location of the local copy can be defined by the user using the software interface.
4. Search inside a temporary directory, which caches single files downloaded from the database before, in order to prevent multiple downloads of the same data. The location of the temporary directory can be defined by the user using the software interface. If desired, the temporary directory can be cleared.
5. Try to download the file from the remote repository. The URL of the remote repository can be defined by the user using the software interface. The downloaded file will be stored inside the temporary directory.

In the third year of Two!Ears further functionalities have been added to the interface allowing for listing and downloading the content of a whole directory with only one command.

1 `urlwrite`, see `http://www.mathworks.fr/help/matlab/ref/urlwrite.html`

3.2 Data Formats

3.2.1 Impulse Responses

For the auralization and rendering of acoustic scenes a-priori measured impulse responses play an important role. The acoustic transmission from a sound source to both ears of the human head can be described by HRIRs or their temporal Fourier transform called Head-Related Transfer Function (HRTF). The measurement of this transfer functions is usually done with a HATS, which models the human outer ears. While HRTFs imply free field conditions and an anechoic environment, BRIRs include reverberation caused by the obstacles and walls. In order to binaurally auralize a sound source emitting a certain stimulus, the respective impulse response is convolved with an anechoic recording of the stimulus. As a standard format for storing and utilizing such impulse responses, the Spatially Oriented Format for Acoustics (SOFA)[11] has been chosen. It represents the basis of the AES-X212 HRTF file format standardization project[12] and is therefore a suitable format for the Two!Ears project.
Besides HRTF and BRIR datasets, MRIRs and multi-loudspeaker BRIRs are supported by the format. MRIRs can be interpreted as a generalization of BRIRs using a multi-channel microphone array instead of a two-channel HATS. While BRIRs can be measured for different source positions by moving the sound source inside the room, multi-loudspeaker BRIRs assume a static setup of loudspeakers and capture the impulse responses for each loudspeaker.

3.2.2 Audio-Visual Scene Description

The definition of a scene description format allows for a formal standardization of audio-visual scenes among all work packages and project partners. It can be interpreted as a physical ground truth and is therefore a useful tool for validating the auditory model against physical labels. It is also used for the configuration of the acoustic rendering tools in order to simulate these scenes appropriately.
During the format's conception binary data containers, e.g, the Hierarchical Data Format (HDF) or the Network Common Data Form (NetCDF), have been considered as a possible basis for the format. They are designed to efficiently handle a large amount data in order to provide flexible access to it. However, special software tools are needed to access and modify the data inside these container files. Hence, the Extensible Markup Language (XML) has been selected as the suitable format for the scene description. It surpasses the mentioned alternatives due to its human readability and cross-platform compatibility. Files written in this format can easily be modified with any text editor. Large binary data, e.g. impulse responses (see Section 3.2.1) or audio stimuli, is referenced inside the scene description file and stored separately.

The structure of the XML files is defined by a so-called XML Schema Definition (XSD), which is utilized for an automatic syntax validity check of the scene description files. An exemplary file is shown in Fig. 3.2. While general rendering parameters are defined as attributes of the `scene` element, scene entities like sound `source`s and `sink`s are characterized within child elements of the `scene`. An audio `buffer` may be added in order to define an input signal for the respective sound source. Basic support of additional visual simulation attributes, e.g. 3D mesh files, is provided. The dynamic part of the scene is described by the `dynamic` element consisting of several `event`s manipulating certain attributes of scene objects. The events' parameters `Start` and `End` can be used to specify a period over which the manipulation is carried out. This is necessary to simulate e.g. continuous motion of a scene object from its current position to the target.

```
<?xml version="1.0" encoding="utf-8"?>
<scene
  BlockSize="4096"
  SampleRate="44100"
  MaximumDelay="0.05"
  PreDelay="0.0"
  LengthOfSimulation="5.0"
  NumberOfThreads="1"
  Renderer="ssr_binaural"
  HRIRs="impulse_responses/qu_kemar_anechoic/QU_KEMAR_anechoic_3m.sofa">
  <source Position="1 2 1.75"
          Type="point"
          Name="Cello"
          MeshFile="cello.mesh"
          Volume="0.4">
    <buffer ChannelMapping="1"
            Type="fifo"
            File="stimuli/anechoic/instruments/anechoic_cello.wav"/>
  </source>
  <source Position="1 -2 1.75"
          Type="point"
          Name="Castanets">
    <buffer ChannelMapping="1"
            Type="fifo"
            File="stimuli/anechoic/instruments/anechoic_castanets.wav"/>
  </source>
  <sink Position="0 0 1.75"
        UnitFront="1 0 0"
        UnitUp="0 0 1"
        Name="Head"/>
  <dynamic>
    <event Name="Castanets"
           Attribute="Position"
           Value="1 2 1.75"
           Start="1.5"
           End="5.0"/>
    <event Name="Cello"
           Attribute="Position"
           Value="1 -2 1.75"
           Start="3.0"/>
  </dynamic>
</scene>
```

Figure 3.2: Exemplary scene description file written in XML

4 Conclusions and Outlook

The synthesis of ear and eye signals plays an important role in the development and evaluation of the Two!Ears model. The simulation framework is fully integrated into the model. The capabilities of the simulation framework were extended during the project. For the ear signals this includes the use of data-based binaural synthesis and static BRIRs for different source and listener positions. The inclusion of dynamic BRIRs remains as an open issue, as it requires a different processing engine compared to the other synthesis methods. The visual simulation framework is extended and integrated into the model.

During the Two!Ears project a database of considerable size has been established. Among others, it includes several utility data to render different acoustic environments and scenarios. In order to evaluate the behaviour of the auditory model against human performance in such environments, a variety of perceptual labels has been added to the database. They cover aspects of the both application cases addressed in Two!Ears, namely Dynamic Auditory Scene Analysis (WP 6.1) and Quality of Experience evaluation (WP 6.2). Many of the datasets have been acquired by members of the consortium. As the database has been made available to community, it states a outstanding development towards reproducible research.

Bibliography

[1] G. Enzner, "Analysis and optimal control of LMS-type adaptive filtering for continuous-azimuth acquisition of head related impulse responses," Las Vegas, NV, April 2008. (Cited on page 11)

[2] N. Hahn and S. Spors, "Measurement of time-variant binaural room impulse responses for data-based synthesis of dynamic auditory scenes," in *German Annual Conference on Acoustics (DAGA)*, March 2014. (Cited on page 12)

[3] ——, "Identification of dynamic acoustic systems by orthogonal expansion of time-variant impulse responses," in *IEEE-EURASIP International Symposium on Control, Communications, and Signal Processing*, May 2014. (Cited on page 12)

[4] S. Spors, H. Wierstorf, and M. Geier, "Comparison of modal versus delay-and-sum beamforming in the context of data-based binaural synthesis," in *132nd Convention of the Audio Engineering Society*, April 2012. (Cited on page 12)

[5] F. Schultz and S. Spors, "Data-based binaural synthesis including rotational and translatory head-movements," in *52nd Conference on Sound Field Control - Engineering and Perception, Audio Engineering Society*, September 2013. (Cited on page 12)

[6] F. Winter, F. Schutz, and S. Spors, "Localization properties of data-based binaural synthesis including translatory head-movements," in *Forum Acousticum*, September 2014, submitted. (Cited on page 13)

[7] T. Lentz, D. Schröder, M. Vorländer, and I. Assenmacher, "Virtual reality system with integrated sound field simulation and reproduction," *EURASIP Journal on Advances in Signal Processing*, vol. 2007, 2007, article ID 70540. (Cited on page 14)

[8] M. Geier and S. Spors, "Spatial Audio Reproduction with the SoundScape Renderer," in *27th Tonmeistertagung - VDT International Convention*, Cologne, Germany, Nov. 2012. [Online]. Available: http://www.int.uni-rostock.de/fileadmin/user_upload/publications/spors/2012/Geier_TMT2012_SSR.pdf (Cited on page 15)

[9] M. Geier, T. Hohn, and S. Spors, "An Open Source C++ Framework for Multithreaded Realtime Multichannel Audio Applications," in *Linux Audio Conference*, Stanford, USA, Apr. 2012. [Online]. Available: http://lac.linuxaudio.org/2012/download/lac2012_proceedings.pdf (Cited on page 15)

[10] J. B. Allen and D. A. Berkley, "Image method for efficiently simulating small-room acoustics," *The Journal of the Acoustical Society of America*, vol. 65, no. 4, 1979. (Cited on page 17)

[11] P. Majdak, Y. Iwaya, T. Carpentier, R. Nicol, M. Parmentier, A. Roginska, Y. Suzuki, K. Watanabe, H. Wierstorf, H. Ziegelwanger, and M. Noisternig, "Spatially Oriented Format for Acoustics: A Data Exchange Format Representing Head-Related Transfer Functions," in *134th Convention of the Audio Engineering Society*, 2013. (Cited on page 22)

[12] M. Parmentier. (2012, Oct.) Project AES-X212, HRTF file format. [Online]. Available: http://www.aes.org/standards/meetings/init-projects/aes-x212-init.cfm (Cited on page 22)

Part II

Content of Public Database

Datasets (Part II)

1 Listening tests **V**

1.1 Human label file format V

1.2 Localisation V

1.2.1 2012-03-01: Localisation of a real vs. binaural simulated point source VI

1.2.2 2013-11-01: Localisation of different source types in sound field synthesis VII

1.2.3 2016-03-11: Localisation of simulatenous talkers by humans and machines X

1.3 Coloration XI

1.3.1 2013-05-01: Coloration of a point source in Wave Field Synthesis . . . XI

1.3.2 2015-10-01: Coloration of a point source in Wave Field Synthesis revisited XIII

1.3.3 2015-10-05: Coloration of a point source in Local Wave Field Synthesis XIV

1.4 Quality ratings XV

1.4.1 2014-04-01: Scene related sound quality XV

1.4.2 2015-11-01: Listening preference of popular music presented by WFS, surround, and stereo XVI

1.4.3 2016-03-01: Listening position preference for different 5.0 reproductions XVIII

1.4.4 2016-06-01: Listening preference of different mixes of one popular music song presented by WFS (binaural simulation) XX

1.4.5 2016-11-18: Listening preference of different mixes of one popular music song presented by WFS XXI

2 Impulse responses **XXIX**

2.1 Anechoic measurements (HRTFs) XXIX

2.1.1 Anechoic HRTFs from the KEMAR manikin with different distances XXIX

2.1.2 Spherical far-field HRTF compilation of the Neumann KU100 . . XXXI

2.1.3 MIT HRTF measurements of a KEMAR dummy head XXXII

2.1.4 Near-field HRTFs from SCUT database of the KEMAR XXXII

2.2 Reverberant measurements (BRIRs) XXXIII

2.2.1 Two!Ears, CNRS Toulouse, Adream-building XXXIII

2.2.2 TU Berlin, room Auditorium 3 XXXV

2.2.3 TU Berlin, room Spirit XXXVII

2.2.4 TU Berlin, room Calypso, 5.0 surround setup for different listening positions XXXVIII

2.2.5 TU Berlin, room Calypso, 19-channel linear loudspeaker array XL

2.2.6 University of Rostock, RIRs and BRIRs of a 64-channel Loudspeaker array for different room configurations . . . XLI
2.2.7 Salford-BBC, 12-channel loudspeaker studio . . . XLIII
2.2.8 University of Surrey, four different rooms . . . XLV
2.2.9 TU Ilmenau, conference room . . . XLVII

3 Trained Models for Knowledge Sources XLIX

4 Sound databases LI
4.1 Speech databases . . . LI
4.1.1 GRID corpus . . . LI
4.2 Acoustic scenes and events . . . LII
4.2.1 IEEE AASP Challenge on Detection and Classification . . . LII

5 Stimuli LV
5.1 Anechoic Stimuli . . . LV
5.1.1 TU Berlin - Noise Stimuli . . . LV
5.1.2 Cologne University of Applied Sciences - Anechoic Recordings . . . LVI
5.1.3 Instruments . . . LVI

6 Visual Stimuli LIX
6.1 Panorama Image of Audio Laboratory at the Institute of Communications Engineering, University of Rostock . . . LIX
6.1.1 License . . . LIX
6.1.2 Description . . . LIX
6.2 Stereo-Vision Capture from Adream Building, CNRS Toulouse . . . LX
6.2.1 License . . . LX
6.2.2 Description . . . LX
6.2.3 Files . . . LX

Bibliography LXIII

1 Listening tests

In this part of the database we collect results from psychoacoustic experiments run by the different labs involved in Two!Ears. In addition to the results we provide the underlying stimuli in a way that they can directly be fed into the Binaural Simulator and provide the model with exactly the same audio input as the listeners experienced during the actual experiments.

1.1 Human label file format

The test results are called human labels in the following and the average results of the single experiments are all stored in the same way in the human label file format. This is a simple csv file, which includes the following entries:

```
# Description of the results stored in the file
# stimuli, rating, 95% confidence interval
experiments/link_to/brs_file1.wav, -0.4653, 0.0123
experiments/link_to/brs_file2.wav,  0.2738, 0.1548
...
```

The file starts with a header that uses # as a comment sign and then includes at least three columns. The first one provides a link to the actual BRS (Binaural Room Scaning) file used in the experiment. The second one the average result from the experiment, this could be a mean, a median, or something else, and the third one showing the variance of the data, in most of the cases in the form of the confidence interval. The files could of course have more columns with additional information, like the time the listeners needed to response or if another value like the sound pressure level was changed during the experiment it could be indicated in a later column.

Besides the human label files, most experiments provide the anonymised results from single listeners. The format of those data can vary, but in all cases they are provided as plain text or csv files.

1.2 Localisation

- *2012-03-01: Localisation of a real vs. binaural simulated point source*
- *2013-11-01: Localisation of different source types in sound field synthesis*
- *2016-03-11: Localisation of simulatenous talkers by humans and machines*

1.2.1 2012-03-01: Localisation of a real vs. binaural simulated point source


Published by members of the Two!Ears consortium

1.2.1.1 Digital Object Identifier

doi: 10.5281/zenodo.164616

1.2.1.2 License

1.2.1.3 Description

In this experiment the localisation of a real point source realised by a loudspeaker was compared to the localisation of a binaural simulation of the same source using HRTFs or BRIRs. The results are published in *[Wierstorf2012]*.

1.2.1.4 Files

The BRS files for every loudspeaker are generated by using the BRIRs from *TU Berlin, room Calypso, 19-channel linear loudspeaker array* and HRTFs from *Anechoic HRTFs from the KEMAR manikin with different distances* are available under:

```
experiments/2012-03-01_brs_vs_real_localisation/brs/*
```

The white noise pulse train under:

```
experiments/2012-03-01_brs_vs_real_localisation/stimulus/white_noise_pulse.wav
```

The results of the 11 listeners for the localisation task and of their head movements while performing this task are available under:

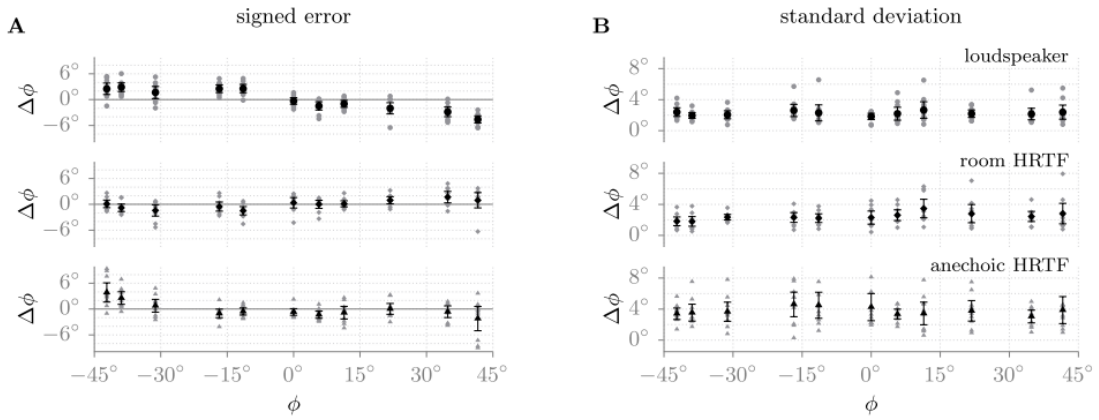


Fig. 1.1: The mean over all subjects together with the 95% confidence interval is shown. In grey, the individual subjects' results are presented. In graph A, the signed error of the localization of the eleven speakers is shown. In graph B, the mean standard deviation for the localization task is depicted. The top row represents the condition with the real loudspeakers, the middle row the room HRTF (Head-Related Transfer Function)s (BRIR (Binaural Room Impulse Response)s), and the bottom row the anechoic HRTFs. Figure from [Wierstorf2012].

```
experiments/2012-03-01_brs_vs_real_localisation/results/*
experiments/2012-03-01_brs_vs_real_localisation/results_head_movements/*
```

An analysis of the results including a plotting script and the average results are available at:

```
experiments/2012-03-01_brs_vs_real_localisation/analysis/*
experiments/2012-03-01_brs_vs_real_localisation/analysis/data_mean/localisation_real_vs_hrir.csv
experiments/2012-03-01_brs_vs_real_localisation/analysis/data_mean/localisation_real_vs_brir.csv
```

1.2.2 2013-11-01: Localisation of different source types in sound field synthesis

Published by members of the Two!Ears consortium

1.2.2.1 Digital Object Identifier

- BRS files: 10.5281/zenodo.55427

- Head movements: 10.5281/zenodo.164620
- Results: 10.5281/zenodo.55439

1.2.2.2 License

Creative Commons Attribution 4.0

1.2.2.3 Description

In this experiment different sound sources like a point source or plane wave were synthesised using WFS (Wave Field Synthesis) and NFC-HOA (Near-Field Compensated Higher Order Ambisonics). In a series of listening test participants had to localise those sources from different listening positions in order to investigate how well the different methods were able to provide the correct directional impression. In order to compare different systems and different listening positions the experiments were performed with dynamic binaural synthesis. The experiment is described in *[Wierstorf2014b]*.

1.2.2.4 Files

The BRS files for the different conditions and the used noise stimulus can be found under:

```
experiments/2013-11-01_sfs_localisation/brs/*
experiments/2013-11-01_sfs_localisation/stimulus/white_noise_pulse.wav
```

The localisation results of the single listeners can be found in the following folders:

```
experiments/2013-11-01_sfs_localisation/results/nfchoa_ps_circular/*
experiments/2013-11-01_sfs_localisation/results/nfchoa_pw_circular/*
experiments/2013-11-01_sfs_localisation/results/wfs_fs_circular/*
experiments/2013-11-01_sfs_localisation/results/wfs_ps_circular/*
experiments/2013-11-01_sfs_localisation/results/wfs_ps_linear/*
experiments/2013-11-01_sfs_localisation/results/wfs_pw_circular/*
```

The results are encoded as follows in the csv files:

trial_number	position in the experiment of the presented condition
mode	0 => real single loudspeaker (for calibration); otherwise number of used loudspeakers (1456 denotes NFC-HOA with 14 loudspeakers, but an order of 28)
source	internal number of presented condition
n	number of head-tracker (always 1)
x	x position of listener (as reported by head-tracker)
y	y-position of listener
z	z-position of listener
phi	horizontal direction of auditory event (includes offset [1])
real_phi	direction of virtual sound event (includes offset [1])
delta	median direction of auditory event [2]
stdev	standard deviation of the listener head orientation during acquisition of head-tracking data (nine samples after the person presses enter)
elapsed_time	time the person needed to localize the source and press enter

In addition, we provide the trajectory of the actual head movements, each listener performed during the localisation experiments in the following folders:

```
experiments/2013-11-01_sfs_localisation/results_head_movements/nfchoa_ps_circular/*
experiments/2013-11-01_sfs_localisation/results_head_movements/nfchoa_pw_circular/*
experiments/2013-11-01_sfs_localisation/results_head_movements/wfs_fs_circular/*
experiments/2013-11-01_sfs_localisation/results_head_movements/wfs_ps_circular/*
experiments/2013-11-01_sfs_localisation/results_head_movements/wfs_ps_linear/*
experiments/2013-11-01_sfs_localisation/results_head_movements/wfs_pw_circular/*
```

Average results can be found at:

```
experiments/2013-11-01_sfs_localisation/analysis/localization_wfs_ps_circular.txt
experiments/2013-11-01_sfs_localisation/analysis/localization_wfs_ps_linear.txt
experiments/2013-11-01_sfs_localisation/analysis/localization_wfs_pw_circular.txt
experiments/2013-11-01_sfs_localisation/analysis/localization_wfs_fs_circular.txt
experiments/2013-11-01_sfs_localisation/analysis/localization_nfchoa_ps_circular.txt
experiments/2013-11-01_sfs_localisation/analysis/localization_nfchoa_pw_circular.txt
```

The average results are encoded in the following way

[1] The offset was varied for the single conditions. Have a look at the average result files for the actual offset values.

[2] The participants were advised to only look into the horizontal plane.

condition	name of used BRS file
X	x-position of listener
Y	y-position of listener
phi	direction of auditory event [3]
phi_error	(direction virtual sound event) - (direction auditory event)
phi_ci	95%-confidence interval of phi
std	standard deviation of the five repetitions used to measure phi
std_ci	95%-confidence interval of std
phi_offset	offset applied to the virtual sound event [4]

1.2.3 2016-03-11: Localisation of simulatenous talkers by humans and machines


Published by members of the Two!Ears consortium

1.2.3.1 License

1.2.3.2 Description

A recent psychophysical study *[KopcoEtAl2010]* has shown that listeners are able to exploit prior knowledge of the masker locations in a cocktail party scenario. This study investigated the ability of listeners to localise a female target voice in the presence of four male masking voices. The is experimental setup is shown in Fig. 1.3. In particular, it addresses two main research questions. First, we ask whether listeners are able to exploit prior information about the masker locations in Kopco et al.'s task when listening over headphones, where binaural cues are limited to those present in the HRIRs used to spatialise the signals for headphone listening. In headphone listening, head movements are not available and room characteristics can be carefully controlled; hence, we also investigate whether prior knowledge of the masker locations can assist localisation in both anechoic and reverberant conditions. The results are shown in Fig. 1.4, where localisation performance for fixed and varied (mixed) masker locations is compared. Second, we ask whether the sources of knowledge available to listeners in this scenario –

[3] If two values are provided in the form of {2,25}, two auditory events were perceived at those two positions.

[4] This offset was introduced to have a jitter for the virtual sound event positions, enabling more randomness to the possible source positions. The results of phi and phi_error are already offset corrected, but not the directions reported in the results files for the single listeners, mentioned above

speaker characteristics and masker locations – can be successfully exploited in a computational system for sound localisation. More details can be found in *[MaBrown2016]*.

1.2.3.3 Files

The results for the individual listeners:

```
experiments/2016-03-11_kopco/results/*
```

An analysis of the results including the resulting plots:

```
experiments/2016-03-11_kopco/analysis/*
experiments/2016-03-11_kopco/analysis/graphics/*
```

1.3 Coloration

- *2013-05-01: Coloration of a point source in Wave Field Synthesis*
- *2015-10-01: Coloration of a point source in Wave Field Synthesis revisited*
- *2015-10-05: Coloration of a point source in Local Wave Field Synthesis*

1.3.1 2013-05-01: Coloration of a point source in Wave Field Synthesis


Published by members of the Two!Ears consortium

1.3.1.1 Digital Object Identifier

10.5281/zenodo.164589

1.3.1.2 License

1.3.1.3 Description

This database entry contains stimuli and results from the experiments described in *[Wierstorf2014a]*. In the experiment different WFS systems synthesising a point source were rated in terms of their perceived coloration compared to a real point source. This was done for different audio material, namely pink noise, speech, and music and different listener positions. The different WFS systems consisted always of a circular loudspeaker array with a radius of 3m, but different number of employed loudspeakers. To control for the exact listening position, allow instantaneous switching between listening positions, and allow for very high numbers of loudspeakers in the WFS systems the experiment was performed with binaural synthesis without head tracking. The results are summarised in Fig. 1.5.

1.3.1.4 Files

The BRS files for the binaural simulation can be found at:

```
experiments/2013-05-01_wfs_coloration/brs/*
```

For the music stimulus a twelve second clip from the electronic song "Luv deluxe" by "Cinnamon Chasers" was chosen, which is not published in this database. The applied speech and pink noise stimuli can be found under:

```
experiments/2013-05-01_wfs_coloration/stimuli/pink_noise_pulse.wav
experiments/2013-05-01_wfs_coloration/stimuli/speech.wav
```

The single results of the listeners and an analysis are available under:

```
experiments/2013-05-01_wfs_coloration/results/*
experiments/2013-05-01_wfs_coloration/analysis/*
```

Where of special interest are the mean rating results shown in Fig. 1.5:

```
experiments/2013-05-01_wfs_coloration/analysis/data_mean/coloration_wfs_center_music.csv
experiments/2013-05-01_wfs_coloration/analysis/data_mean/coloration_wfs_center_noise.csv
experiments/2013-05-01_wfs_coloration/analysis/data_mean/coloration_wfs_center_speech.csv
experiments/2013-05-01_wfs_coloration/analysis/data_mean/coloration_wfs_multiple_music.csv
experiments/2013-05-01_wfs_coloration/analysis/data_mean/coloration_wfs_multiple_noise.csv
experiments/2013-05-01_wfs_coloration/analysis/data_mean/coloration_wfs_multiple_speech.csv
```

Note: This experiment has a flaw for the results using very high number of loudspeakers. In this case some of the perceivable coloration was due to some artifacts arriving

from the fact that we employed only integer delay in the delay line involved in the WFS processing. See the next experiment for a corrected version.

1.3.2 2015-10-01: Coloration of a point source in Wave Field Synthesis revisited


Published by members of the Two!Ears consortium

1.3.2.1 Digital Object Identifier

10.5281/zenodo.164592

1.3.2.2 License

1.3.2.3 Description

The first experiment investigating coloration in WFS had some artefacts in the stimuli for high frequencies due to the limitation coming with a finite sampling rate during the process of time delaying single signals in WFS. This time the experiment was repeated using fractional delay filter. In addition, we added a linear loudspeaker array besides the circular one used in the first experiment.

1.3.2.4 Files

The BRS files for the binaural simulation, and the speech and pink noise stimuli are available under:

```
experiments/2015-10-01_wfs_coloration/brs/*
experiments/2015-10-01_wfs_coloration/stimuli/pink_noise_pulse.wav
experiments/2015-10-01_wfs_coloration/stimuli/speech.wav
```

The results of the single listeners and an analysis are available under:

```
experiments/2015-10-01_wfs_coloration/results/*
experiments/2015-10-01_wfs_coloration/analysis/*
```

The average results, plotted in Fig. 1.6, are stored under:

```
experiments/2015-10-01_wfs_coloration/analysis/data_mean/coloration_wfs_circular_center_music.csv
experiments/2015-10-01_wfs_coloration/analysis/data_mean/coloration_wfs_circular_center_noise.csv
experiments/2015-10-01_wfs_coloration/analysis/data_mean/coloration_wfs_circular_center_speech.csv
experiments/2015-10-01_wfs_coloration/analysis/data_mean/coloration_wfs_circular_offcenter_music.csv
experiments/2015-10-01_wfs_coloration/analysis/data_mean/coloration_wfs_circular_offcenter_speech.csv
experiments/2015-10-01_wfs_coloration/analysis/data_mean/coloration_wfs_linear_center_music.csv
experiments/2015-10-01_wfs_coloration/analysis/data_mean/coloration_wfs_linear_center_noise.csv
experiments/2015-10-01_wfs_coloration/analysis/data_mean/coloration_wfs_linear_center_speech.csv
experiments/2015-10-01_wfs_coloration/analysis/data_mean/coloration_wfs_linear_offcenter_music.csv
experiments/2015-10-01_wfs_coloration/analysis/data_mean/coloration_wfs_linear_offcenter_speech.csv
```

1.3.3 2015-10-05: Coloration of a point source in Local Wave Field Synthesis


Published by members of the Two!Ears consortium

1.3.3.1 License

1.3.3.2 Description

In the experiment a circular 56-channel loudspeaker array with a radius of 3 m was driven by different sound field synthesis techniques in order to reproduce a virtual point source. The different techniques were WFS, spectral band-limited NFC-HOA, and local WFS, which is similar to spectral band-limited NFC-HOA in showing a high aliasing frequency in a small region. The techniques were rated in terms of their perceived coloration compared to a real point source using a Multi-Stimulus with Hidden Anchor and Reference (MUSHRA) test paradigm. This was done for different audio material, namely pink noise, speech, and music. The listener was positioned in the center of the loudspeaker array. The experiments has been repeated for an off center listening position. To control for the exact listening position, allow instantaneous switching between listening positions, and the experiment was performed with binaural synthesis without head tracking. The results are summarised in Fig. 1.7 and Fig. 1.8.

1.3.3.3 Files

The BRS files for the binaural simulation are available under:

```
experiments/2015-10-05_localwfs_coloration/brs/*
```

The results of the single listeners and an analysis are available under:

```
experiments/2015-10-01_wfs_coloration/results/*
experiments/2015-10-01_wfs_coloration/analysis/*
```

The average results as well as the statistical significances plotted in the figures, are stored under:

```
experiments/2015-10-05_localwfs_coloration/analysis/stats_music_center.txt
experiments/2015-10-05_localwfs_coloration/analysis/stats_music_off.txt
experiments/2015-10-05_localwfs_coloration/analysis/stats_noise_center.txt
experiments/2015-10-05_localwfs_coloration/analysis/stats_noise_off.txt
experiments/2015-10-05_localwfs_coloration/analysis/stats_speech_center.txt
experiments/2015-10-05_localwfs_coloration/analysis/stats_speech_off.txt
```

Note: This experiment has a flaw that the used sample rate of the stimuli and for creating the BRS files was 48000 Hz while the sample rate for the HRTF dataset was 44100 Hz. Hence, the HRTF were scaled with respect to frequency.

1.4 Quality ratings

- *2014-04-01: Scene related sound quality*
- *2015-11-01: Listening preference of popular music presented by* WFS*, surround, and stereo*
- *2016-03-01: Listening position preference for different 5.0 reproductions*
- *2016-06-01: Listening preference of different mixes of one popular music song presented by* WFS *(binaural simulation)*
- *2016-11-18: Listening preference of different mixes of one popular music song presented by* WFS

1.4.1 2014-04-01: Scene related sound quality


Published by members of the Two!Ears consortium

1.4.1.1 Digital Object Identifier

doi: 10.5281/zenodo.164624

1.4.1.2 License

1.4.1.3 Description

We did a paired comparison preference test where we asked for the preferred audio quality of the presented stimuli, see *[Raake2014]* for details. The stimuli consisted always of the same acoustic scene with three individual sources, which had different degrees of impairments: no impairment (gray), medium degraded guitar (light red), degraded guitar (red), degraded vocals (blue). Those impairments were applied in different combinations to the three sources, see Fig. 1.9. The ordering in Fig. 1.9 also corresponds to the average listening preference for the presented scenes, where 1. is the most preferred scene and 10. the less preferred one.

1.4.1.4 Files

We provide the ratings of the listeners and the wav-files of the stimuli in the following folders. The wav-files were created by using the HRTFs from *Anechoic HRTFs from the KEMAR manikin with different distances*:

```
experiments/2014-04-01_scene_quality/stimuli/*
experiments/2014-04-01_scene_quality/results/*
```

The result files contain always three columns, in the first two columns is the number of the presented condition (the number corresponds to the number at the beginning of the stimuli files). The third row contains a `1` or `2` indicated which of the two presented conditions was preferred by the listeners.

1.4.2 2015-11-01: Listening preference of popular music presented by WFS, surround, and stereo


Published by members of the Two!Ears consortium

1.4.2.1 Digital Object Identifier

- Stimuli: doi: 10.14279/depositonce-5173
- Results: doi: 10.5281/zenodo.164433

1.4.2.2 License

- Stimuli: Creative Commons Attribution-NonCommercial-ShareAlike 4.0
- Results: Creative Commons Attribution 4.0

1.4.2.3 Description

We did a paired comparison preference test where listeners rated their listening preference for four different pop musical pieces presented by WFS, stereo or surround. The musical pieces were all mixed by the same person in order to try to minimize the influence of the mix on the ratings, but still trying to get the best out of every system, see *[Hold2016a]* for details. As this experiment was performed with real loudspeakers, there are no BRS files available with this experiment, but if you like to run a binaural model on the stimuli you can try the BRS files from *2016-06-01: Listening preference of different mixes of one popular music song presented by WFS (binaural simulation)* which provide a anechoic binaural simulation of the loudspeaker array setup used in this experiment.

1.4.2.4 Files

```
experiments/2015-11-01_wfs_stereo_comparison/results/*
experiments/2015-11-01_wfs_stereo_comparison/stimuli/brew-stereo.wav
experiments/2015-11-01_wfs_stereo_comparison/stimuli/brew-surround.wav
experiments/2015-11-01_wfs_stereo_comparison/stimuli/brew-wfs.wav
experiments/2015-11-01_wfs_stereo_comparison/stimuli/brew2-stereo.wav
experiments/2015-11-01_wfs_stereo_comparison/stimuli/brew2-surround.wav
experiments/2015-11-01_wfs_stereo_comparison/stimuli/brew2-wfs.wav
experiments/2015-11-01_wfs_stereo_comparison/stimuli/lighthouse-stereo.wav
experiments/2015-11-01_wfs_stereo_comparison/stimuli/lighthouse-surround.wav
experiments/2015-11-01_wfs_stereo_comparison/stimuli/lighthouse-wfs.wav
experiments/2015-11-01_wfs_stereo_comparison/stimuli/lighthouse2-stereo.wav
experiments/2015-11-01_wfs_stereo_comparison/stimuli/lighthouse2-surround.wav
experiments/2015-11-01_wfs_stereo_comparison/stimuli/lighthouse2-wfs.wav
experiments/2015-11-01_wfs_stereo_comparison/stimuli/sister-stereo.wav
experiments/2015-11-01_wfs_stereo_comparison/stimuli/sister-surround.wav
experiments/2015-11-01_wfs_stereo_comparison/stimuli/sister-wfs.wav
experiments/2015-11-01_wfs_stereo_comparison/stimuli/toynbee-stereo.wav
```

```
experiments/2015-11-01_wfs_stereo_comparison/stimuli/toynbee-surround.wav
experiments/2015-11-01_wfs_stereo_comparison/stimuli/toynbee-wfs.wav
```

1.4.3 2016-03-01: Listening position preference for different 5.0 reproductions


Published by members of the Two!Ears consortium

1.4.3.1 Digital Object Identifier

doi-10.5281/zenodo.164413

1.4.3.2 License

Creative Commons Attribution 4.0

1.4.3.3 Description

We conducted an experiment on the preferred listening position in a 5.0 surround setup. The experiment was performed in the studio like room Calypso in the Telefunken building of the TU Berlin. The experiment employed dynamic binaural synthesis in order to allow instantaneous switching between positions during listening. To accomplish this BRIRs were recorded before at nine different positions, see *TU Berlin, room Calypso, 5.0 surround setup for different listening positions*. 26 test participants rated their preferred listening position out of the 9 positions for every recording technique. They did this first without visual feedback using a GUI (Graphical User Interface) that only showed nine buttons where the stimuli were randomly assigned to. In a second run they had a GUI showing a actual sketch of the listening setup, similar to the sketch shown in Table 2.3.

Note: The BRIRs used in this experiment were not the one available under 10.5281/zenodo.160761, but an earlier version available under 10.5281/zenodo.49691. That earlier version had an error in the metadata of the stored SOFA files, resulting in the fact that the listeners were not looking towards the front at all listening positions, but facing always towards the center loudspeaker.

As music material seven different simultaneously recordings of the piece “Maurerische Trauermusik K.477” of W. A. Mozart were used. The recordings differed in the applied

recording technique, which are listed in Table 1.1 and were done at ORF (Austrian Broadcast, Vienna) and the piece was played by the Radio Symphony Orchestra Vienna. For more details on the applied microphones and for downloading all recordings have a look at *[Wittek2015]*.

Fig. 1.11 summarizes the results across all listeners and recording techniques, showing only differences between with and without visual feedback. The results for the different recording techniques are summarized in this PDF. For more details on the experiment have a look at *[Schultze2016]*.

Table 1.1: Different recording techniques used and there corresponding abbreviations.

Abbreviation	Recording technique
Rec. 80	Stereo + C + NHK
Rec. 81	Decca-Tree + NHK
Rec. 82	OCT + NHK
Rec. 83	INA5 (Brauner ASM5)
Rec. 84	Schoeps KFM 360 + DSP-4 KFM 360
Rec. 85	OCT-Surround
Rec. 86	Soundfield MKV + SP 451

1.4.3.4 Files

The experiment was performed with dynamic binaural synthesis. The employed BRS files are available under:

```
experiments/2016-03-01_sweet_spot/brs/
```

The actual stimuli were extracted from the DVD available for download at *[Wittek2015]* and are stored in the database as wav files under:

```
experiments/2016-03-01_sweet_spot/stimuli/
```

The results of the single listeners are available under:

```
experiments/2016-03-01_sweet_spot/results/no_visual_feedback/
experiments/2016-03-01_sweet_spot/results/visual_feedback/
```

The analysis, resulting in Fig. 1.11 is available under:

```
experiments/2016-03-01_sweet_spot/analysis/
```

1.4.4 2016-06-01: Listening preference of different mixes of one popular music song presented by WFS (binaural simulation)

 Published by members of the Two!Ears consortium

1.4.4.1 Digital Object Identifier

- Stimuli: doi: 10.5281/zenodo.61000
- Results: doi: 10.5281/zenodo.162161

1.4.4.2 License

1.4.4.3 Description

We did a paired comparison preference test where listeners rated their listening preference for a pop musical pieces presented by WFS or stereo. For WFS four to five different mixes of the same song were presented for one run of the experiment. Those mixes were variations along one of the following mixing parameters: “compression”, “equalizing”, “reverb”, “changes of reverb, equalizing and compression to the vocals alone”, “positioning of musical foreground elements”. The musical pieces were all mixed by the same person, that also mixed the stereo and WFS reference condition, compare *[Hold2016a]*. The results for the “positioning of musical foreground elements” are discussed in more detail in *[Hold2016b]*.

1.4.4.4 Files

The experiment was performed with dynamic binaural synthesis. The BRS files for the 56-channel circular loudspeaker array are available under:

```
experiments/2016-06-01_wfs_mixing_quality/brs/
```

The actual stimuli for the different mixing conditions are available as wav file, containing the driving signals for the loudspeakers. The stereo file is available as a two channel signal, the WFS signals are stored in a 64-channel wav file, where the first 56 channel are the driving signals for the loudspeakers:

```
experiments/2016-06-01_wfs_mixing_quality/stimuli/
```

XX

The single preference ratings of the 41 test participants and the results of the survey conducted after the experiment are available under:

```
experiments/2016-06-01_wfs_mixing_quality/results/
```

The data analysis scripts and results, that calculates the preference values shown in Fig. 1.12 are available under:

```
experiments/2016-06-01_wfs_mixing_quality/analysis/
```

1.4.5 2016-11-18: Listening preference of different mixes of one popular music song presented by WFS

 Published by members of the Two!Ears consortium

1.4.5.1 Digital Object Identifier

- Stimuli: doi: 10.5281/zenodo.61000
- Results: doi: 10.5281/zenodo.167331

1.4.5.2 License

Creative Commons Attribution 4.0

1.4.5.3 Description

We did a paired comparison preference test where listeners rated their listening preference for a pop musical pieces presented by WFS or stereo. For WFS four to five different mixes of the same song were presented for one run of the experiment. Those mixes were variations along one of the following mixing parameters: “compression”, “equalizing”, “reverb”, “changes of reverb, equalizing and compression to the vocals alone”, “positioning of musical foreground elements”. The musical pieces were all mixed by the same person, that also mixed the stereo and WFS reference condition, compare *[Hold2016a]*. This experiment is very similar to an *earlier one* with the main difference that the earlier one used binaural synthesis to present the stimuli, whereby here we have used real loudspeaker setups.

1.4.5.4 Files

The actual stimuli for the different mixing conditions are available as wav file, containing the driving signals for the loudspeakers. The stereo file is available as a two channel signal, the WFS signals are stored in a 64-channel wav file, where the first 56 channel are the driving signals for the loudspeakers:

```
experiments/2016-11-18_wfs_mixing_quality/stimuli/
```

The single preference ratings of the 41 test participants and the results of the survey conducted after the experiment are available under:

```
experiments/2016-11-18_wfs_mixing_quality/results/
```

The data analysis scripts and results, that calculates the preference values shown in Fig. 1.13 are available under:

```
experiments/2016-11-18_wfs_mixing_quality/analysis/
```

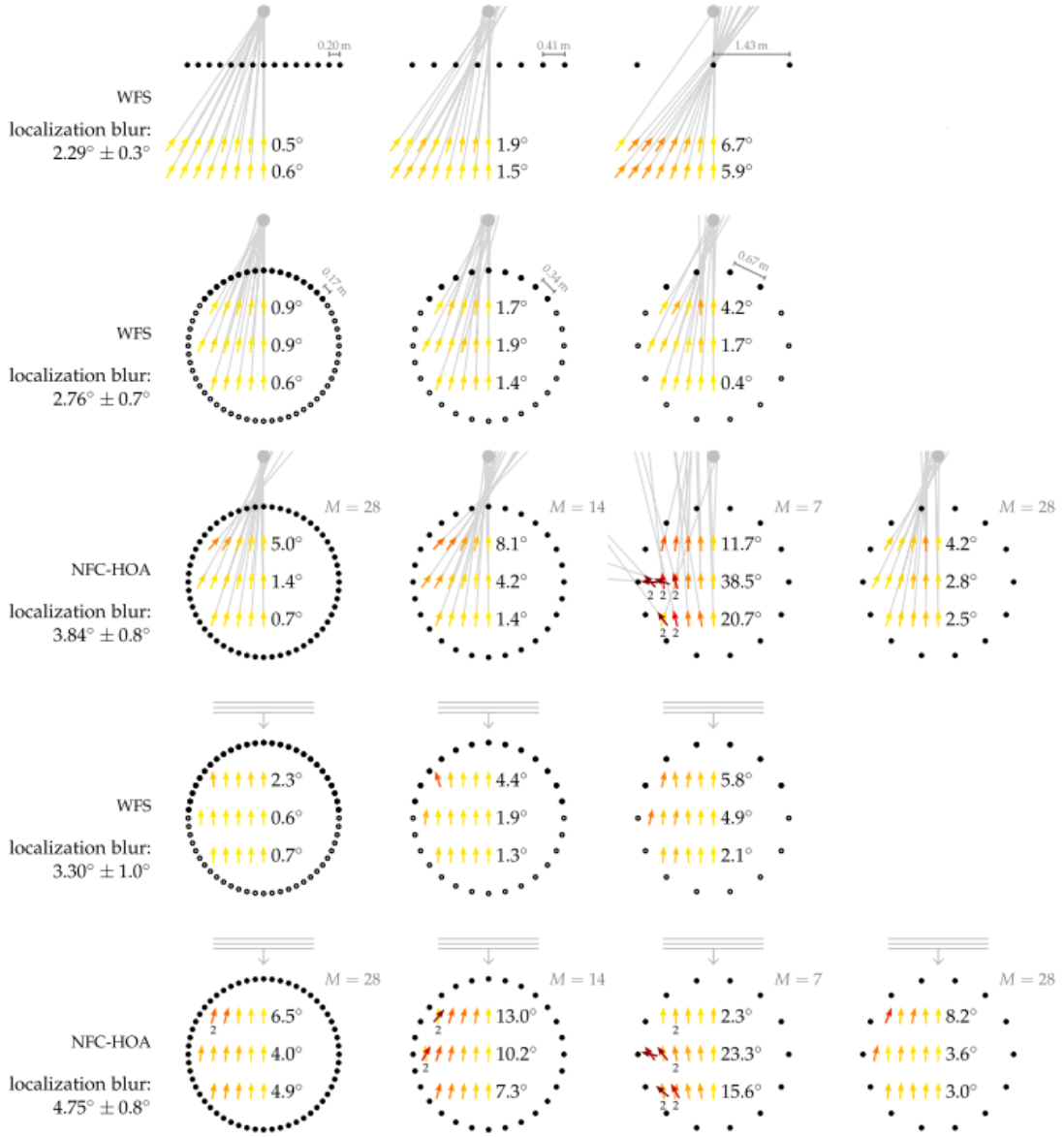


Fig. 1.2: Average localization results. The black symbols indicate loudspeakers, the grey ones the synthesised source. At every listening position, an arrow is pointing into the direction from which the listeners perceived the corresponding auditory event. The color of the arrow displays the absolute localization error, which is also summarised as an average beside the arrows for every row of positions. The average confidence interval for all localization results is 2.3°. Listening conditions which resulted in listeners saying that they perceived two sources are highlighted with a small 2 written below the position. Figure from [Wierstorf2014b].

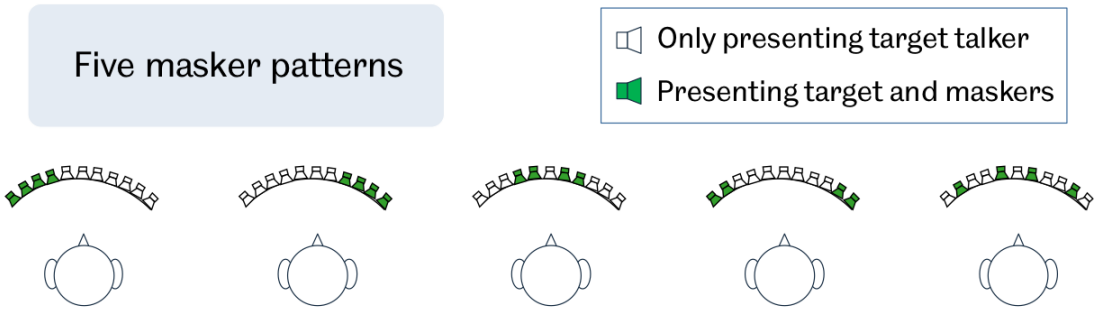


Fig. 1.3: Experimental settings used in this study. One target female voice and four interfering male voices are presented at the same time from 11 potential loudspeaker locations. Five masker patterns are considered as indicated by the green loudspeakers.

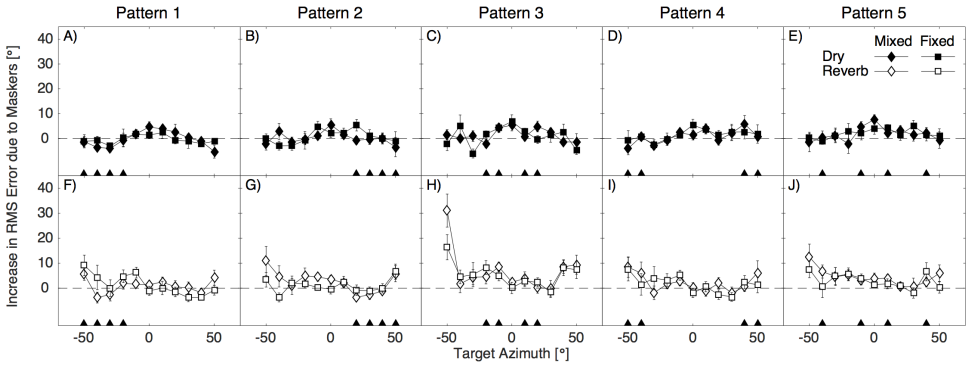


Fig. 1.4: Across-participant average (± standard error of the mean) of the increases in root-mean-square errors (with respect to the no-masker control condition) as a function of the target location. Masker locations are indicated by the filled triangles along the abscissa. Each column of panels shows the Mixed and Fixed condition data for one masker pattern and for the anechoic session (upper panels) and the reverberant session (lower panels).

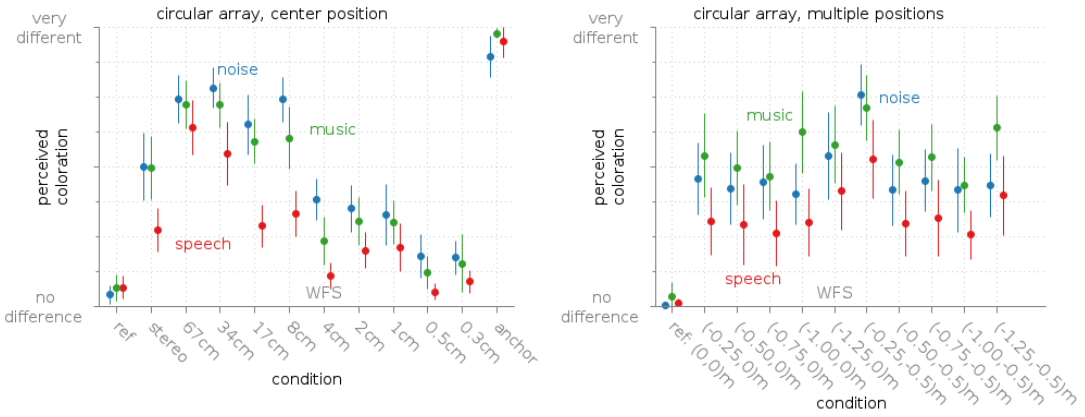


Fig. 1.5: Mean rated coloration of a point source synthesised with WFS compared to a reference point source. The error bars are showing the confidence intervals.

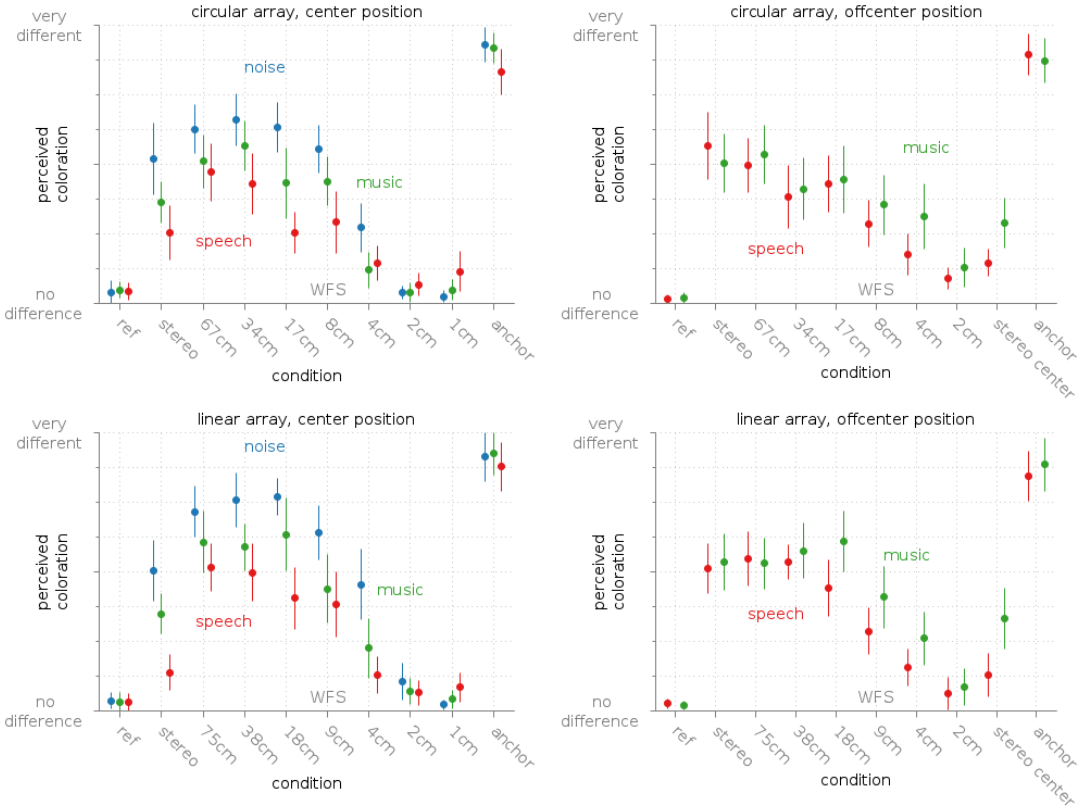


Fig. 1.6: Median of the rated coloration of a point source synthesized with WFS compared to a reference point source. The error bars are showing the confidence intervals.

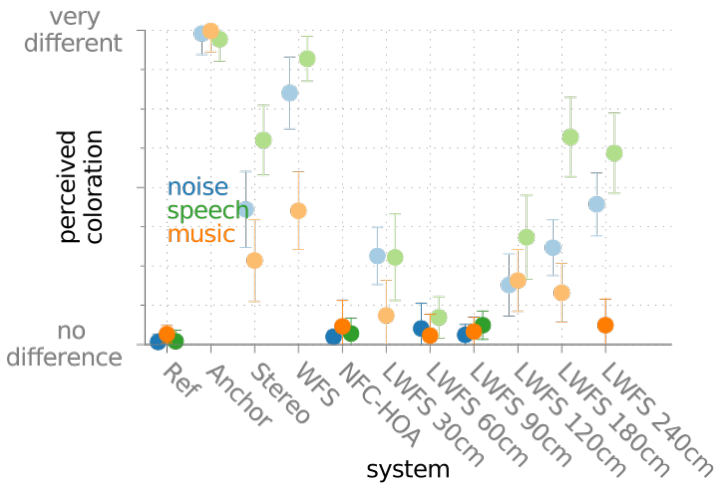


Fig. 1.7: Median of the rated coloration of a point source synthesized with different sound field synthesis techniques compared to a reference point source for a centered listening positions. The error bars are showing the first and the third quartile. The points with lower saturated color show a significant difference from the reference. The length given after Local Wave Field Synthesis (LWFS) denotes the radius of the local listening area. For details see the `PDF version` of this figure.

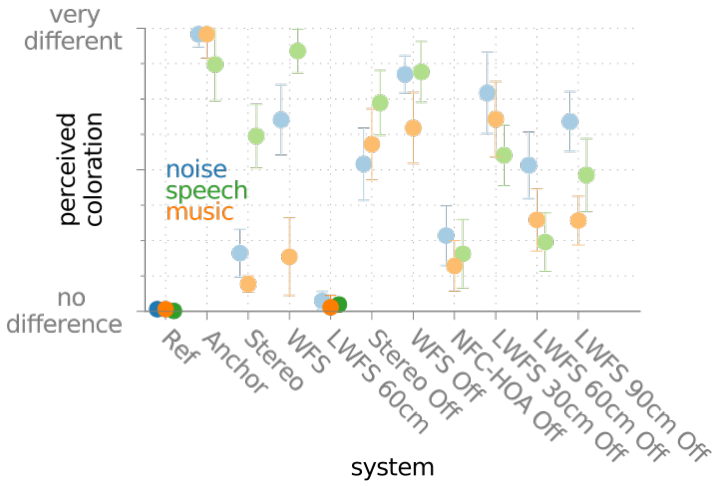


Fig. 1.8: Median of the rated coloration of a point source synthesized with different sound field synthesis techniques compared to a reference point source for an off-center listening position (marked by Off). For details see the `PDF version` of this figure.

Fig. 1.9: The ten different conditions of the quality test, ordered after their listener preferrence. See text for the meaning of the color code.

Fig. 1.10: Estimated preference of the different reproduction systems averaged over all listeners and musical pieces. The preference was estimated with a Bradley-Terry-Luce model. The bars denote the corresponding 95% confidence interval. `(PDF version)`

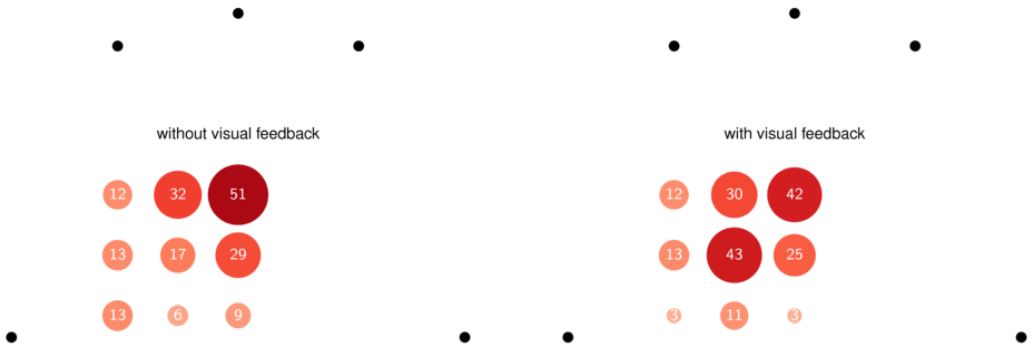


Fig. 1.11: Number of chosen preferred positons summed over all listeners and recording techniques. On the left side the results without visual feedback and on the right with visual feedback about the actual listening position are shown. `(PDF version)`

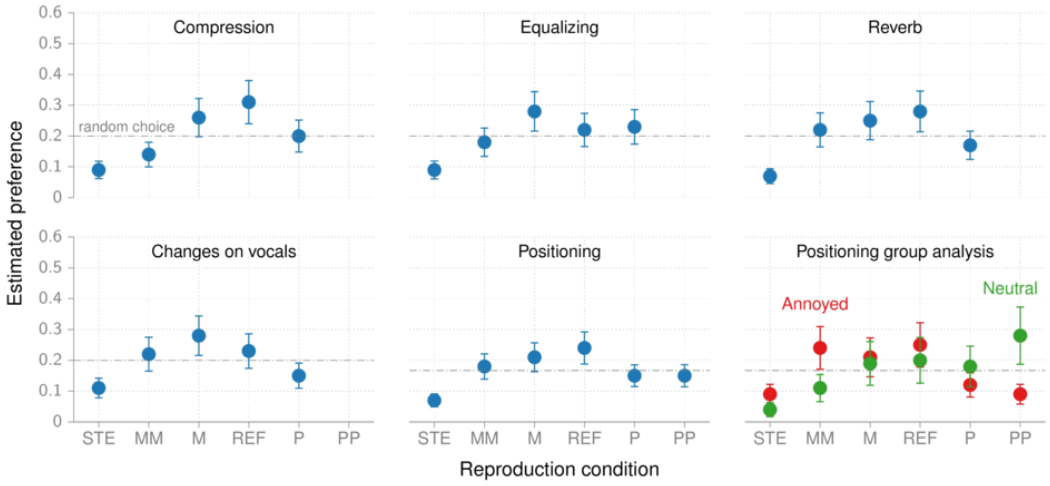


Fig. 1.12: Estimated preferences of the different mixes of a piece of pop music for WFS. The conditions are stereo (STE), WFS reference (REF), both from *a former listening test*, WFS with mixing parameter switched off (MM), WFS with decreased processing of the specified mixing parameter (M), WFS with increased processing (P), and WFS with further increased processing (PP). The preference was estimated with a Bradley-Terry-Luce model. The bars denote the corresponding 95% confidence interval. `(PDF version)`

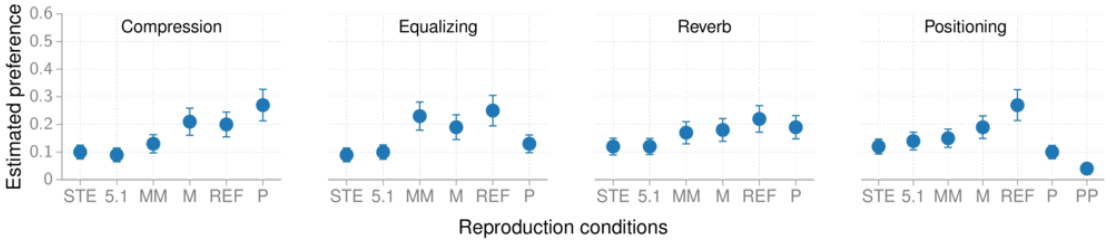


Fig. 1.13: Estimated preferences of the different mixes of a piece of pop music for WFS. The conditions are stereo (STE), surround (5.1) WFS reference (REF), both from *a former listening test*, WFS with mixing parameter switched off (MM), WFS with decreased processing of the specified mixing parameter (M), WFS with increased processing (P), and WFS with further increased processing (PP). The preference was estimated with a Bradley-Terry-Luce model. The bars denote the corresponding 95% confidence interval. `(PDF version)`

2 Impulse responses

The Two!Ears database provides a large collection of HRTFs and BRIRs. Some of those were measured by the Two!Ears project, others were collected from other labs. All of them are available in the SOFA format which allows for an easy usage with the Binaural Simulator.

2.1 Anechoic measurements (HRTFs)

- *Anechoic* HRTF*s from the KEMAR manikin with different distances*
- *Spherical far-field* HRTF *compilation of the Neumann KU100*
- *MIT* HRTF *measurements of a KEMAR dummy head*
- *Near-field HRTFs from SCUT database of the KEMAR*

2.1.1 Anechoic HRTFs from the KEMAR manikin with different distances


Published by members of the Two!Ears consortium

2.1.1.1 Digital Object Identifier

doi: 10.5281/zenodo.55418

2.1.1.2 License

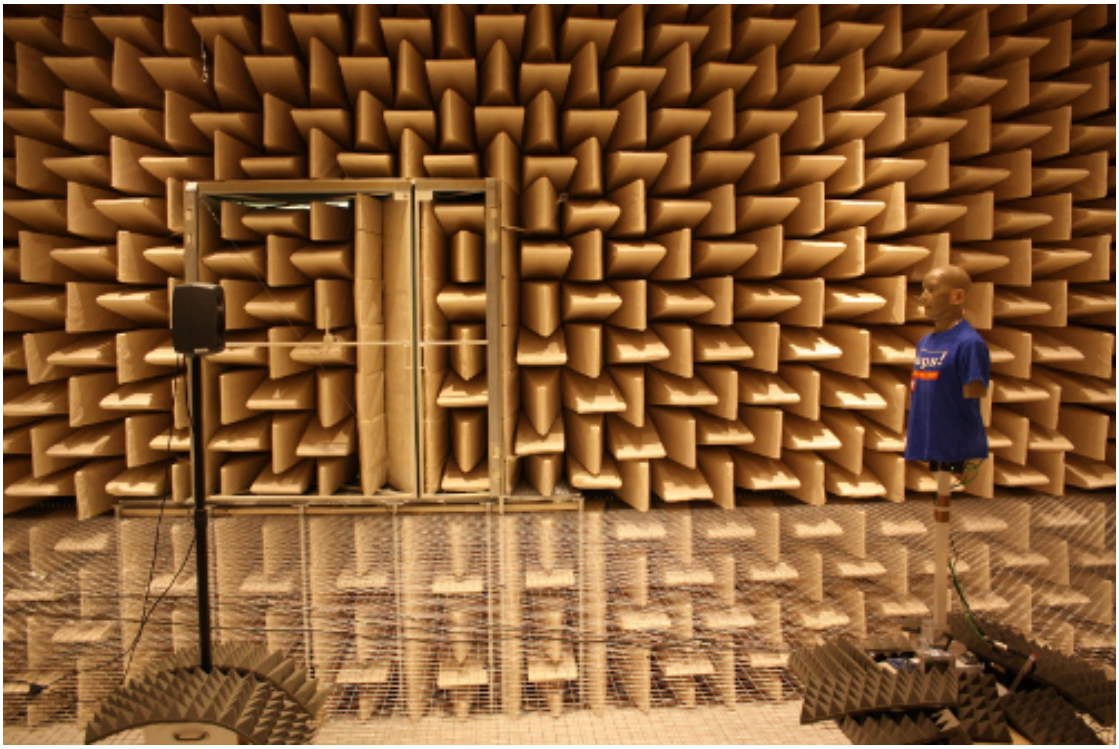


Fig. 2.1: Setup of the KEMAR in the anechoic chamber of TU Berlin.

2.1.1.3 Description

HRTFs measured with a KEMAR dummy head in the anechoic chamber of the TU Berlin *[Wierstorf2011]*. The HRTFs were measured in the horizontal plane with a resolution of 1° for the three different distances of 0.5m, 1m, 2m, 3m.

Note, that for the distance of 0.5m the used Genelec loudspeaker presents not really a point source. In addition, the HRTFs for 0.5m include reflections of the sound going from the KEMAR head back to the loudspeaker and back to the dummy head.

2.1.1.4 Files

```
impulse_responses/qu_kemar_anechoic/QU_KEMAR_anechoic_0_5m.sofa
impulse_responses/qu_kemar_anechoic/QU_KEMAR_anechoic_1m.sofa
impulse_responses/qu_kemar_anechoic/QU_KEMAR_anechoic_2m.sofa
impulse_responses/qu_kemar_anechoic/QU_KEMAR_anechoic_3m.sofa
```

XXX

The measurement comes also with the following headphone compensation filters:

```
impulse_responses/qu_kemar_anechoic/QU_KEMAR_AKGK271_hcomp.wav
impulse_responses/qu_kemar_anechoic/QU_KEMAR_AKGK601_hcomp.wav
impulse_responses/qu_kemar_anechoic/QU_KEMAR_SennheiserHD25_hcomp.wav
```

2.1.2 Spherical far-field HRTF compilation of the Neumann KU100

2.1.2.1 License

Creative Commons Attribution-ShareAlike 3.0

2.1.2.2 Description

Fig. 2.2: Setup of the measurement in the anechoic chamber of FH Köln.

Three-dimensional HRIR datasets were measured with the Neumann KU100 dummy head. An active 3-way louspeaker (Genelec 8260A) was used as a sound source with a constant distance of approximately 3.25m. Different apparent source positions were realized by rotating the dummy head around two axis using the VariSphear measurement system *[Bernschuetz2010]*. The impulse responses were captured for different sampling configurations of the source's position:

- horizontal plane with a resolution of 1°
- two different equidistant spherical Lebedev grids with 2354 and 2702 sampling points
- full sphere equiangular 2° Gauss grid with 16020 sampling points

For further details, see the FH Köln website or the corresponding paper *[Bernschuetz2013]*.

2.1.2.3 Files

```
impulse_responses/fhk_ku100_anechoic/HRIR_CIRC360RM.sofa
impulse_responses/fhk_ku100_anechoic/HRIR_CIRC360.sofa
impulse_responses/fhk_ku100_anechoic/HRIR_FULL2DEG.sofa
```

2.1.3 MIT HRTF measurements of a KEMAR dummy head

2.1.3.1 Description

The three-dimensional HRTF dataset was measured with a KEMAR (type DB-4004) equipped with a large right ear (type DB-065) and a normal-size left ear (type DB-061). A small two-way loudspeaker (Realistic Optimus Pro 7) was used as a sound source. The HRTFs were measured for a distances of 1.4m. The elevation angle varies from -40° (40° below horizontal plane) to +90° (directly overhead) with a stepsize of 10°. The azimuth angle varies from 0° to 360° with an elevation angle dependent resolution. Files were downloaded from the SOFA database. For documentation see *[Gardner1994]* which is available here.

2.1.3.2 Files

```
impulse_responses/mit_kemar_anechoic/MIT_KEMAR_anechoic_1.7m_large.sofa
impulse_responses/mit_kemar_anechoic/MIT_KEMAR_anechoic_1.7m_normal.sofa
```

2.1.4 Near-field HRTFs from SCUT database of the KEMAR

2.1.4.1 Description

The three-dimensional HRTF dataset was measured with a KEMAR dummy head. The HRTFs were measured for ten different distances of 0.2m, 0.25m, 0.3m and 0.4m, 0.5m, ..., 1.0m. The elevation angle varies from -30° to +90° with a stepsize of 15°. The azimuth angle varies from 0° to 360° with a resolution of 5° for elevation angles between ±30°. Above +30° elevation angle the azimuthal resolution is 10° , while for +90° elevation only one measurement per distance was performed. Files were downloaded from the SOFA database. See *[Xie2013a]* and *[Xie2013b]* for documentation on the measurements.

2.1.4.2 Files

```
impulse_responses/scut_kemar_anechoic/SCUT_KEMAR_anechoic.sofa
impulse_responses/scut_kemar_anechoic/SCUT_KEMAR_anechoic_0.2m.sofa
impulse_responses/scut_kemar_anechoic/SCUT_KEMAR_anechoic_0.3m.sofa
impulse_responses/scut_kemar_anechoic/SCUT_KEMAR_anechoic_0.4m.sofa
impulse_responses/scut_kemar_anechoic/SCUT_KEMAR_anechoic_0.5m.sofa
impulse_responses/scut_kemar_anechoic/SCUT_KEMAR_anechoic_0.6m.sofa
impulse_responses/scut_kemar_anechoic/SCUT_KEMAR_anechoic_0.7m.sofa
impulse_responses/scut_kemar_anechoic/SCUT_KEMAR_anechoic_0.8m.sofa
impulse_responses/scut_kemar_anechoic/SCUT_KEMAR_anechoic_0.9m.sofa
impulse_responses/scut_kemar_anechoic/SCUT_KEMAR_anechoic_0.25m.sofa
impulse_responses/scut_kemar_anechoic/SCUT_KEMAR_anechoic_1m.sofa
```

2.2 Reverberant measurements (BRIRs)

- *Two!Ears, CNRS Toulouse, Adream-building*
- *TU Berlin, room Auditorium 3*
- *TU Berlin, room Spirit*
- *TU Berlin, room Calypso, 5.0 surround setup for different listening positions*
- *TU Berlin, room Calypso, 19-channel linear loudspeaker array*
- *University of Rostock,* RIR *(Room Impulse Response)s and* BRIR*s of a 64-channel Loudspeaker array for different room configurations*
- *Salford-BBC, 12-channel loudspeaker studio*
- *University of Surrey, four different rooms*
- *TU Ilmenau, conference room*

2.2.1 Two!Ears, CNRS Toulouse, Adream-building


Published by members of the Two!Ears consortium

2.2.1.1 Digital Object Identifier

doi: 10.5281/zenodo.49357

2.2.1.2 License

2.2.1.3 Description

The Adream-building at the University Toulouse is a robot-lab space consisting of a large hall that contains four small rooms without concrete windows, doors, and ceilings. Fig. 2.3 provides a look from above on the part where the measurements took place.

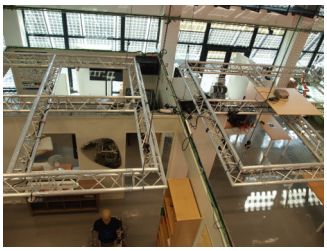


Fig. 2.3: Lab space in the Adream building, look from above.

Fix listener positions with head movements

The measurement consists of two parts. The first part contains BRIRs with head rotations from -78° to 78° in 2° steps, as from the listeners perspective. Those measurements were done at four different listener positions and consisted each time of four different loudspeakers. Fig. 2.4 summarizes the setup with the actual loudspeaker and dummy head positions and orientations.

Moving listener without head rotations

The second part of the measurement consists of a trajectory of 20 listener positions at which BRIRs were measured. This measurement includes only a fixed head orientation of 0°, but the same 4 loudspeaker positions as the first one. Fig. 2.5 summarizes the setup with the actual loudspeaker and dummy head positions and orientations.

2.2.1.4 Files

```
impulse_responses/twoears_kemar_adream/TWOEARS_KEMAR_ADREAM_pos1.sofa
impulse_responses/twoears_kemar_adream/TWOEARS_KEMAR_ADREAM_pos2.sofa
impulse_responses/twoears_kemar_adream/TWOEARS_KEMAR_ADREAM_pos3.sofa
impulse_responses/twoears_kemar_adream/TWOEARS_KEMAR_ADREAM_pos4.sofa
impulse_responses/twoears_kemar_adream/TWOEARS_KEMAR_ADREAM_trajectory.sofa
```

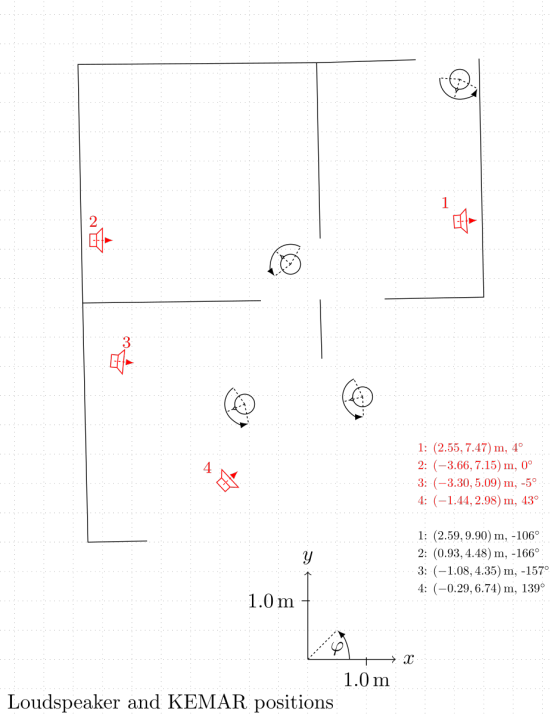


Fig. 2.4: Setup of the measurements. For details see the PDF version of this figure.

2.2.2 TU Berlin, room Auditorium 3

Published by members of the Two!Ears consortium

2.2.2.1 Digital Object Identifier

doi: 10.5281/zenodo.160749

2.2.2.2 License

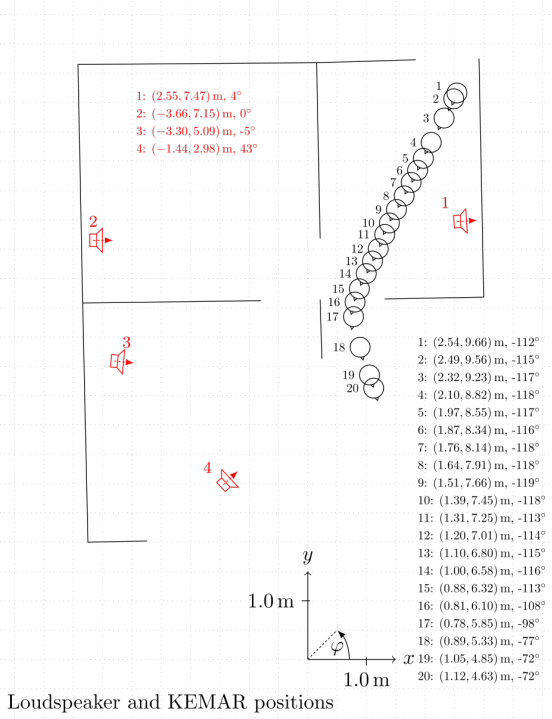
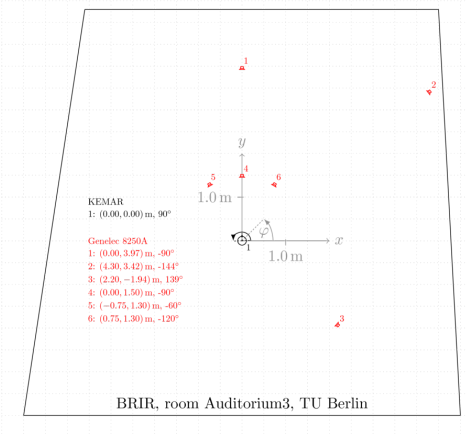


Fig. 2.5: Setup of the measurements. For details see the PDF version of this figure.

2.2.2.3 Description

Table 2.1: Measurement setup. Click on the figures for larger versions.

Sketch of setup (PDF version)	Photo from the position of loudspeaker 1
	

The BRIRs were measured at the mid-size lecture room Auditorium 3 at the Telefunken-

building of TU Berlin. They were measured for six different loudspeaker positions. The head of the dummy head was rotated with a resolution of 1° ranging from -90° to 90°. The measurement equipment was the same as described in *[Wierstorf2011]*.

2.2.2.4 Files

```
impulse_responses/qu_kemar_rooms/auditorium3/QU_KEMAR_Auditorium3.sofa
```

2.2.3 TU Berlin, room Spirit

Published by members of the Two!Ears consortium

2.2.3.1 Digital Object Identifier

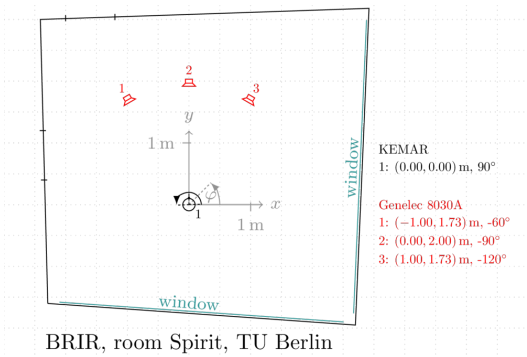
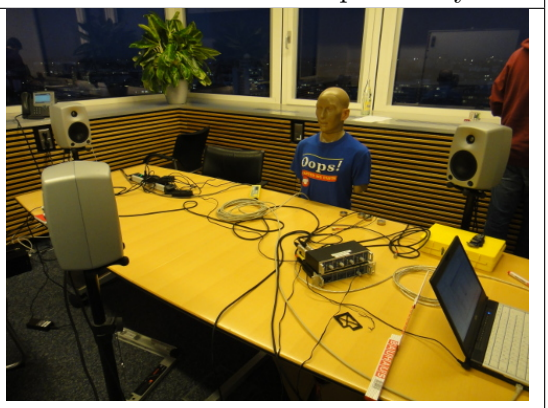
doi-10.5281-zenodo.160751

2.2.3.2 License

Creative Commons Attribution 4.0

2.2.3.3 Description

Table 2.2: Measurement setup. Click on the figures for larger versions.

Sketch of setup (PDF version)	Photo with different loudspeaker layout
 BRIR, room Spirit, TU Berlin	

The BRIRs were measured in the meeting room Spirit at the Telefunken-building of TU Berlin. They were measured for three different sources and with a resolution of 1° and head movement from -90° to 90°. Note, that the photo of the room was not taken for the actual measurement setup. The measurement equipment was the same as described in *[Wierstorf2011]*.

2.2.3.4 Files

```
impulse_responses/qu_kemar_rooms/spirit/QU_KEMAR_spirit.sofa
```

2.2.4 TU Berlin, room Calypso, 5.0 surround setup for different listening positions


Published by members of the Two!Ears consortium

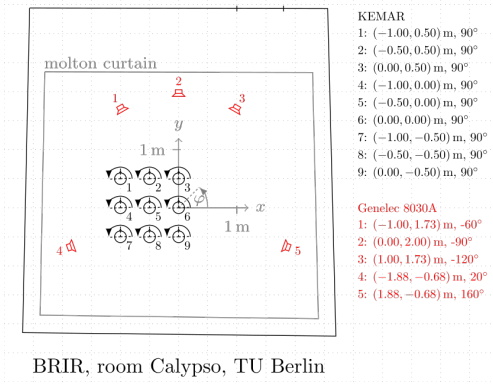
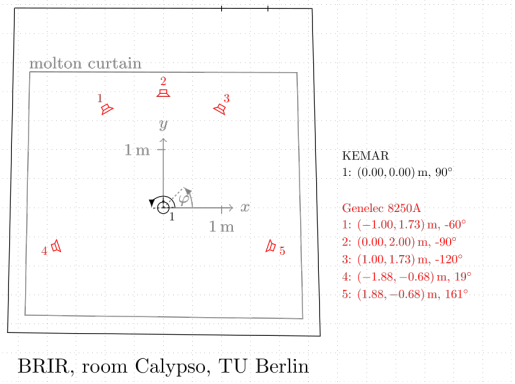
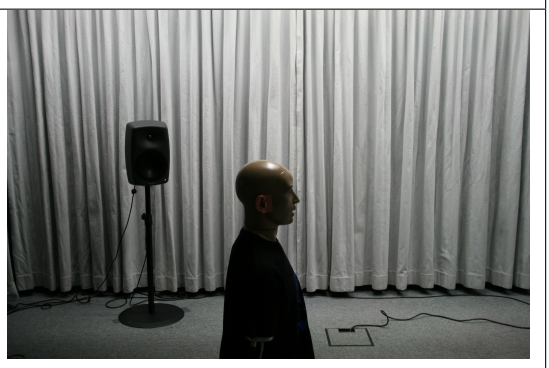
2.2.4.1 Digital Object Identifier

doi: 10.5281/zenodo.160761

2.2.4.2 License

2.2.4.3 Description

Table 2.3: Sketches (top) and photographs (bottom) of measurement setup.

Genelec 8030A (PDF version)	Genelec 8250A (PDF version)
 BRIR, room Calypso, TU Berlin	 BRIR, room Calypso, TU Berlin
	

The BRIRs were measured using a KEMAR (type 45BA) with the corresponding large ears (type KB0065 and KB0066) in the studio room Calypso at the Telefunken-building of TU Berlin. 5 Loudspeakers were placed around the manikin to establish a 5.0 surround setup. The measurements were done with two different two-way loudspeaker fabricates, namely Genelec 8030A (left figures) and Genelec 8250A (right). For the former, the measurements were repeated at 9 different listening positions (bottom left). For each position of the manikin, its head was rotated horizontally from -90° to 90° with a resolution of 1°. The room has a volume of 83 m^3 and a reverberation time RT60 of 0.17 s at a frequency of 1 kHz.

2.2.4.4 Files

BRIRs for Genelec 8030A:

```
impulse_responses/qu_kemar_rooms/calypso_surround/genelec8030A/1_KEMAR_Calypso_Surround_X-1.0m_Y+0.5m.sofa
impulse_responses/qu_kemar_rooms/calypso_surround/genelec8030A/2_KEMAR_Calypso_Surround_X-0.5m_Y+0.5m.sofa
```

```
impulse_responses/qu_kemar_rooms/calypso_surround/genelec8030A/3_KEMAR_Calypso_Surround_X+0.0m_Y+0.5m.sofa
impulse_responses/qu_kemar_rooms/calypso_surround/genelec8030A/4_KEMAR_Calypso_Surround_X-1.0m_Y+0.0m.sofa
impulse_responses/qu_kemar_rooms/calypso_surround/genelec8030A/5_KEMAR_Calypso_Surround_X-0.5m_Y+0.0m.sofa
impulse_responses/qu_kemar_rooms/calypso_surround/genelec8030A/6_KEMAR_Calypso_Surround_X+0.0m_Y+0.0m.sofa
impulse_responses/qu_kemar_rooms/calypso_surround/genelec8030A/7_KEMAR_Calypso_Surround_X-1.0m_Y-0.5m.sofa
impulse_responses/qu_kemar_rooms/calypso_surround/genelec8030A/8_KEMAR_Calypso_Surround_X-0.5m_Y-0.5m.sofa
impulse_responses/qu_kemar_rooms/calypso_surround/genelec8030A/9_KEMAR_Calypso_Surround_X+0.0m_Y-0.5m.sofa
impulse_responses/qu_kemar_rooms/calypso_surround/genelec8030A/KEMAR_Calypso_Surround.sofa
```

BRIRs for Genelec 8250A:

```
impulse_responses/qu_kemar_rooms/calypso_surround/genelec8250A/KEMAR_Calypso_Surround_Large_X+0.0m_Y+0.0m.sofa
```

2.2.5 TU Berlin, room Calypso, 19-channel linear loudspeaker array


Published by members of the Two!Ears consortium

2.2.5.1 Digital Object Identifier

doi: 10.5281/zenodo.160754

2.2.5.2 License

2.2.5.3 Description

BRIRs for 19 different loudspeakers placed in room Calypso at the Telefunken-building of TU Berlin were measured. The room is a studio listening room. The 19 loudspeakers constituted a linear loudspeaker array with a inter-loudspeaker distance of roughly 15cm. The measurement was done with the KEMAR (type 45BA) with the corresponding large ears (type KB0065 and KB0066) and Fostex PM0.4 loudspeakers. The dummy head was rotated from -90° to 90° in 1° steps. The measurement was repeated with the head wearing AKG K601 open headphones.

2.2.5.4 Files

Fig. 2.6: KEMAR wearing AKG K601 headphones in front of the loudspeaker array hidden by a acoustically transparent curtain.

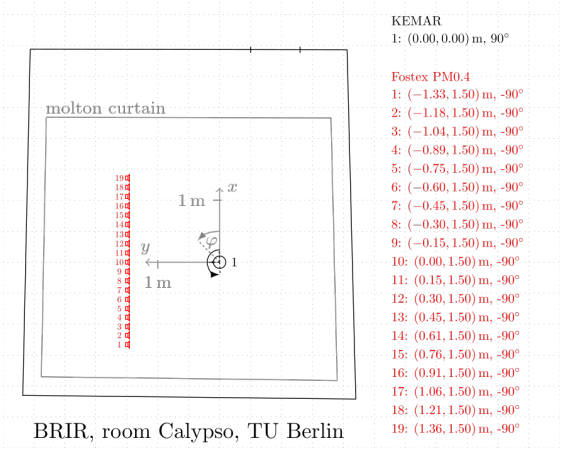


Fig. 2.7: Setup of the measurements. For details see the `PDF version` of this figure.

```
qu_kemar_rooms/calypso_loudspeaker_array/QU_KEMAR_Calypso_loudspeaker_array.sofa
qu_kemar_rooms/calypso_loudspeaker_array/QU_KEMAR_Calypso_loudspeaker_array_AKG_K601.sofa
```

2.2.6 University of Rostock, RIRs and BRIRs of a 64-channel Loudspeaker array for different room configurations

2.2.6.1 Digital Object Identifier

doi: 10.142799/depositonce-87.6

2.2.6.2 License

Creative Commons Attribution 4.0

2.2.6.3 Description

The database contains measured single-channel and binaural room impulse responses (RIRs and BRIRs) of a 64-channel loudspeaker array of square shape under varying room acoustical conditions. The measurements have been performed at the Audio Lab of the Institute of Communications Engineering, University of Rostock. The RIRs have been measured at three receiver positions for four different absorber configurations. Corresponding BRIRs for head-orientations in the range of $\pm 80°$ in $2°$ steps with a KEMAR manikin have been captured for a subset of seven combinations of position and absorber configurations.

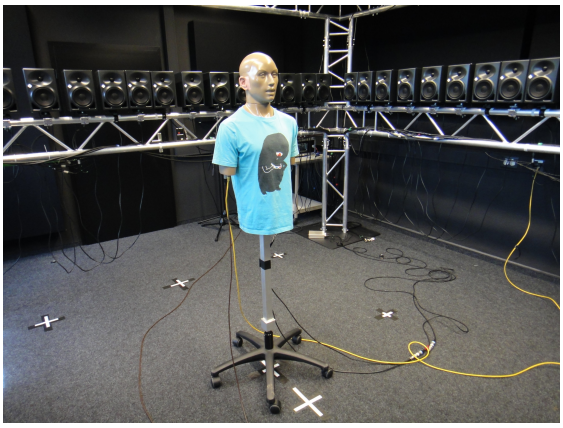


Fig. 2.8: KEMAR Manikin in Audio Laboratory.

2.2.6.4 Files

Binaural room impulse responses:

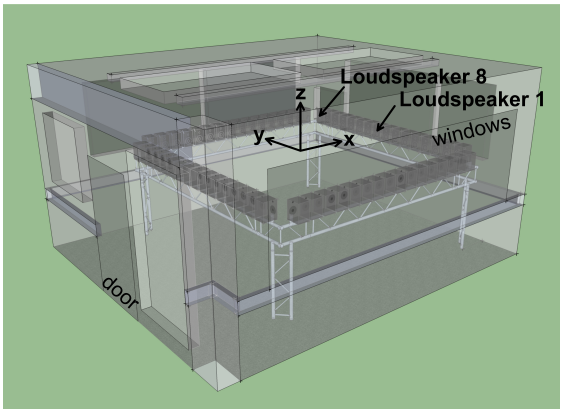


Fig. 2.9: Sketch of Audio Laboratory.

```
impulse_responses/uro_kemar_audiolab/brirs/BRIR_AddAbsorbers_ArrayCentre_Emitters1to64.sofa
impulse_responses/uro_kemar_audiolab/brirs/BRIR_NoAbsorbers_ArrayCentre_Emitters1to64.sofa
impulse_responses/uro_kemar_audiolab/brirs/BRIR_AllAbsorbers_ArrayCentre_Emitters1to64.sofa
impulse_responses/uro_kemar_audiolab/brirs/BRIR_NoAbsorbers_OffCentre_Emitters1to64.sofa
impulse_responses/uro_kemar_audiolab/brirs/BRIR_AllAbsorbers_OffCentre_Emitters1to64.sofa
impulse_responses/uro_kemar_audiolab/brirs/BRIR_NoCeilingAbsorbers_ArrayCentre_Emitters1to64_shifted.sofa
impulse_responses/uro_kemar_audiolab/brirs/BRIR_AllAbsorbers_RoomCentre_Emitters1to64.sofa
```

Single-channel impulse responses:

```
impulse_responses/uro_kemar_audiolab/rirs/RIR_AddAbsorbers_ArrayCentre_Emitters1to64.sofa
impulse_responses/uro_kemar_audiolab/rirs/RIR_AddAbsorbers_OffCentre_Emitters1to64.sofa
impulse_responses/uro_kemar_audiolab/rirs/RIR_AddAbsorbers_RoomCentre_Emitters1to64.sofa
impulse_responses/uro_kemar_audiolab/rirs/RIR_AllAbsorbers_ArrayCentre_Emitters1to64.sofa
impulse_responses/uro_kemar_audiolab/rirs/RIR_AllAbsorbers_OffCentre_Emitters1to64.sofa
impulse_responses/uro_kemar_audiolab/rirs/RIR_AllAbsorbers_RoomCentre_Emitters1to64.sofa
impulse_responses/uro_kemar_audiolab/rirs/RIR_NoAbsorbers_ArrayCentre_Emitters1to64.sofa
impulse_responses/uro_kemar_audiolab/rirs/RIR_NoAbsorbers_OffCentre_Emitters1to64.sofa
impulse_responses/uro_kemar_audiolab/rirs/RIR_NoAbsorbers_RoomCentre_Emitters1to64.sofa
impulse_responses/uro_kemar_audiolab/rirs/RIR_NoCeilingAbsorbers_ArrayCentre_Emitters1to64.sofa
impulse_responses/uro_kemar_audiolab/rirs/RIR_NoCeilingAbsorbers_OffCentre_Emitters1to64.sofa
impulse_responses/uro_kemar_audiolab/rirs/RIR_NoCeilingAbsorbers_RoomCentre_Emitters1to64.sofa
```

2.2.7 Salford-BBC, 12-channel loudspeaker studio

2.2.7.1 License

2.2.7.2 Description

The BRIRs were measured for a 12-channel Genelec 8030A loudspeaker setup inside the University of Salford's ITU-R BS.1116-compliant listening room. The B&K Type 4100 head-and-torso-simulator (HATS) was positioned at 15 different positions. The torso of the HATS was rotated horizontally in the range of ±180° in 2° steps at each position. Additional information can be found at the authors' website.

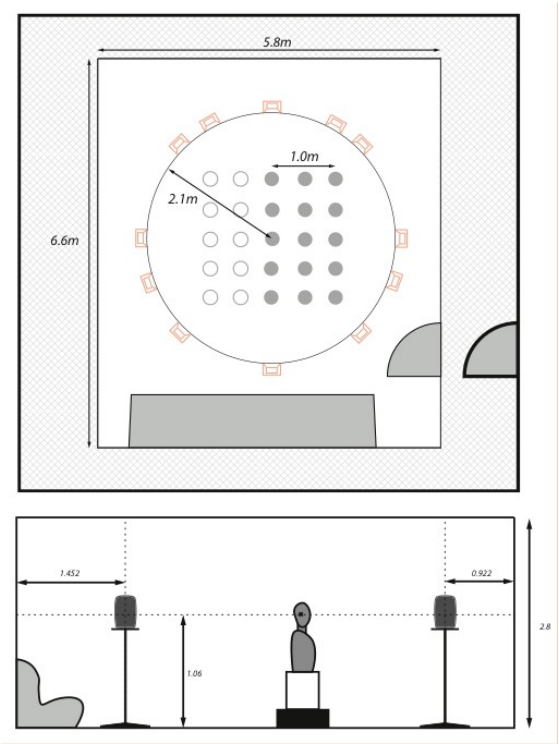


Fig. 2.10: Measurement layout.

2.2.7.3 Files

```
impulse_responses/bbc_bk_salford/SBSBRIR_x-0pt5y-0pt5.sofa
impulse_responses/bbc_bk_salford/SBSBRIR_x0pt5y-0pt5.sofa
impulse_responses/bbc_bk_salford/SBSBRIR_x-0pt5y0.sofa
impulse_responses/bbc_bk_salford/SBSBRIR_x0pt5y0.sofa
impulse_responses/bbc_bk_salford/SBSBRIR_x-0pt5y-1.sofa
impulse_responses/bbc_bk_salford/SBSBRIR_x0pt5y-1.sofa
impulse_responses/bbc_bk_salford/SBSBRIR_x0y-0pt5.sofa
impulse_responses/bbc_bk_salford/SBSBRIR_x0y0.sofa
impulse_responses/bbc_bk_salford/SBSBRIR_x0y-1.sofa
```

```
impulse_responses/bbc_bk_salford/SBSBRIR_x-1y-0pt5.sofa
impulse_responses/bbc_bk_salford/SBSBRIR_x1y-0pt5.sofa
impulse_responses/bbc_bk_salford/SBSBRIR_x-1y0.sofa
impulse_responses/bbc_bk_salford/SBSBRIR_x1y0.sofa
impulse_responses/bbc_bk_salford/SBSBRIR_x-1y-1.sofa
impulse_responses/bbc_bk_salford/SBSBRIR_x1y-1.sofa
```

2.2.8 University of Surrey, four different rooms

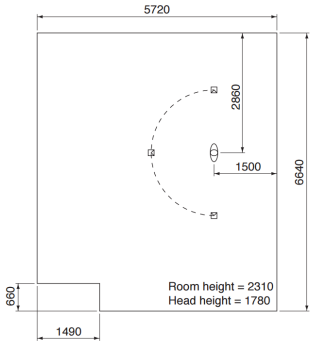
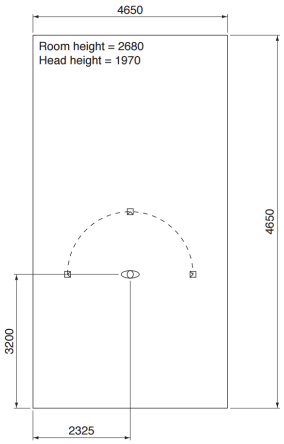
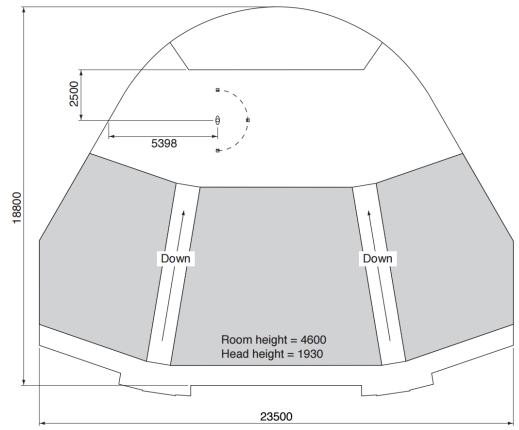
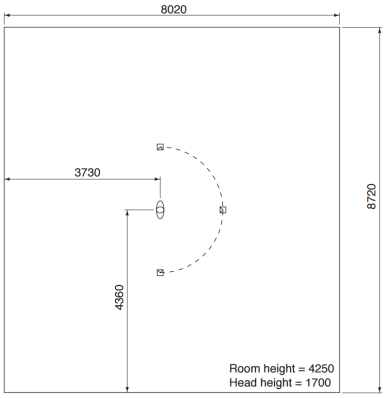
2.2.8.1 License

2.2.8.2 Description

The BRIRs are measured with the Cortex Instruments Mk.2 HATS in four different rooms. A Genelec 8020A active loudspeaker has been used as the sound source. While position and orientation of the HATS is kept constant for one room, the sound source was placed in the horizontal plane at 1.5m distance with an azimuth between ±90° with an increment of 5°. Additional information can be found at the author's website, on the github page and in *[Hummersone2011]*.

The Data has been converted from the original SimpleFreeFieldHRIR convention to the SingleRoomDRIR convention (see also here). Furthermore the azimuth angle convention was been changed from clockwise counting to counter-clockwise counting.

Table 2.4: Rooms of Dataset [Hummersone2011]

A: medium sized office	B: medium-small class room
	
D: large cinema	**C: medium-large seminar room**
	

2.2.8.3 Files

```
impulse_responses/surrey_cortex_rooms/UniS_Room_A_BRIR_16k.sofa
impulse_responses/surrey_cortex_rooms/UniS_Room_A_BRIR_48k.sofa
impulse_responses/surrey_cortex_rooms/UniS_Room_B_BRIR_16k.sofa
impulse_responses/surrey_cortex_rooms/UniS_Room_B_BRIR_48k.sofa
impulse_responses/surrey_cortex_rooms/UniS_Room_C_BRIR_16k.sofa
impulse_responses/surrey_cortex_rooms/UniS_Room_C_BRIR_48k.sofa
impulse_responses/surrey_cortex_rooms/UniS_Room_D_BRIR_16k.sofa
impulse_responses/surrey_cortex_rooms/UniS_Room_D_BRIR_48k.sofa
```

2.2.9 TU Ilmenau, conference room

2.2.9.1 Digital Object Identifier

doi: 10.5281/zenodo.163617

2.2.9.2 License

2.2.9.3 Description

The BRIRs were measured with a KEMAR 45BA head-and-torso-simulator (HATS) with large ears using an active 2-way loudspeaker Genelec 1030A as the sound source. The HATS was positioned at five different locations, while the sound source has always been 2.5 meter in front of the HATS. The head of the HATS was rotated horizontally in the range of ±180° in 5° steps at each position. Additional information can be found in the corresponding paper *[Neidhardt2016]*.

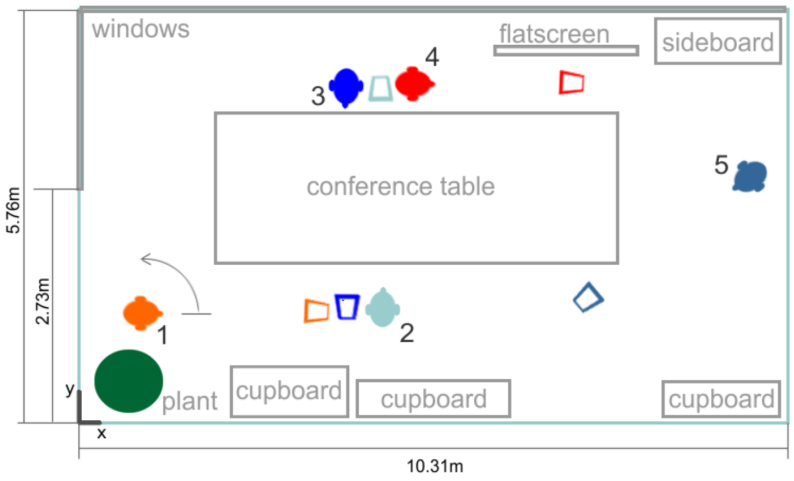


Fig. 2.11: Measurement layout. For details see the `PDF version` of this figure.

2.2.9.4 Files

```
impulse_responses/tuil_kemar_confroom/confroom_pos1.sofa
impulse_responses/tuil_kemar_confroom/confroom_pos2.sofa
impulse_responses/tuil_kemar_confroom/confroom_pos3.sofa
impulse_responses/tuil_kemar_confroom/confroom_pos4.sofa
impulse_responses/tuil_kemar_confroom/confroom_pos5.sofa
```

3 Trained Models for Knowledge Sources

In order to ease up the usage of Two!Ears model, the training of the propabilistic models used for the knowledge sources may be skipped by the user. This is a collection of data created during a-priori training of the models. The data for the respective knowledge source is stored inside `learned_models/<name of knowledge source class>`. Documentation of the training data is available inside the description of the individual sec-knowledge-sources.

L

4 Sound databases

This is a collection of different sound databases. A sound database can be anything that comes as a collection of related sound files like a speech corpus.

4.1 Speech databases

- *GRID corpus*

4.1.1 GRID corpus

GRID is a large multi-talker audiovisual sentence corpus to support joint computational-behavioural studies in speech perception. In brief, the corpus consists of high-quality audio and video (facial) recordings of 1000 sentences spoken by each of 34 talkers (18 male, 16 female). Sentences are of the form:

```
put red at G9 now
```

The database provided here is a subset of the original database containing 360 randomly selected sentences for each speaker. More details about GRID can be found on the GRID website or in the corresponding GRID paper.

4.1.1.1 License

The GRID corpus, together with transcriptions, is freely available for research use.

4.1.1.2 Usage

Each speaker comes with separate folder containing 360 sentences:

```
sound_databases/grid_subset/s1/bbaf2n.wav
...
sound_databases/grid_subset/s1/swwv9a.wav
sound_databases/grid_subset/s2/bbaf1n.wav
...
sound_databases/grid_subset/s34/swws3n.wav
```

All available files are listed in:

```
sound_databases/grid_subset/flist.txt
```

4.2 Acoustic scenes and events

- *IEEE AASP Challenge on Detection and Classification*

4.2.1 IEEE AASP Challenge on Detection and Classification

This data set includes stereo recordings of acoustic environmental scenes as well as isolated events (of an office environment). It can be used for classification tasks of acoustic scenes and events.

We have put the isolated sounds into a folder structure processable for the sec-identification-training, removed the printer class, as it was not very suited for the training, and added a "void" class with different kinds of noise (to be used as negative examples during training).

4.2.1.1 License

Creative Commons Attribution 2.0 UK: England & Wales

4.2.1.2 Usage

To use the data base for acoustic event classification, stereo WAV files from the following folders are of interest:

LII

```
alert/
clearthroat/
cough/
doorslam/
drawer/
keyboard/
keys/
knock/
laughter/
mouse/
pageturn/
pendrop/
phone/
speech/
switch/
void/
```

In each folder different recordings of the corresponding class are provided along with an annotation file with on- and offset times.

In the `scenes/` folder stereo recordings of the following acoustic scenes are provided:

bus
busystreet (busy street with heavy traffic)
office
openairmarket (open-air or semi-open market)
park
quietstreet (quiet street with mild traffic)
restaurant
supermarket
tube (train in the Transport for London, Underground and Overground, train networks)
tubestation (tube station in the Transport for London, Underground and Overground, train networks, either subterranean or supraterranean)

Each class contains ten recordings. Each recording is 30 s long. Files are named according to the class name, i.e. `classXX.wav` where `XX` is a two-digit, non-consecutive number.

LIV

5 Stimuli

5.1 Anechoic Stimuli

5.1.1 TU Berlin - Noise Stimuli

5.1.1.1 License

CC0, Public Domain Dedication

5.1.1.2 Description

A 100 s train of Gaussian white noise pulses with a duration of 700 ms and a pause of 300 ms between each pulse is provided. The single pulses are independent white noise signals. They were windowed with a Hanning window of 20 ms length at their start and end. The resulting signal was band-pass filtered with a fourth order Butterworth filter between 125 Hz and 20000 Hz.

In addition, a 5.2 s train of pulsed pink noise with a duration of 800ms and a pause of 500 ms between each pulse is provided. Here, the single pulses are identical pink noise signals. They were windowed with a Hanning window of 50 ms length at their start and end.

5.1.1.3 Files

```
stimuli/anechoic/aipa/pink_noise_pulse.wav
stimuli/anechoic/aipa/white_noise_pulse.wav
```

5.1.2 Cologne University of Applied Sciences - Anechoic Recordings

5.1.2.1 License

5.1.2.2 Description

The recordings of two Flamengo and two Pop/Blues pieces were done in the anechoic chamber of the Cologne University of Applied Sciences, Institute of Communication Engineering. The sampling rate of all files was changed from 48000 to 44100 Hertz. The original files are available at the university's website.

5.1.2.3 Files

```
stimuli/anechoic/fh-koeln/BluesA_GitL_44100.wav
stimuli/anechoic/fh-koeln/BluesA_GitR_44100.wav
stimuli/anechoic/fh-koeln/BluesA_Voc_44100.wav
stimuli/anechoic/fh-koeln/Flamenco1_U89_44100.wav
stimuli/anechoic/fh-koeln/Flamenco2_U89_44100.wav
stimuli/anechoic/fh-koeln/HellOfAGuy_GitL_44100.wav
stimuli/anechoic/fh-koeln/HellOfAGuy_GitR_44100.wav
stimuli/anechoic/fh-koeln/HellOfAGuy_Voc_44100.wav
```

5.1.3 Instruments

5.1.3.1 Description

Tiny collection of anechoic instrument recordings containing a castanet and a cello.

5.1.3.2 Files

```
stimuli/anechoic/instruments/castanets.wav
stimuli/anechoic/instruments/cello.wav
```

LVI

This is a collection of anechoic stimuli like noise or cello that you can use to convolve with the *impulse responses*. Those files are placed in the *Anechoic Stimuli* section.

6 Visual Stimuli

6.1 Panorama Image of Audio Laboratory at the Institute of Communications Engineering, University of Rostock

6.1.1 License

CC0 1.0 Universal

6.1.2 Description

Cylindrical 360° indoor panorama `vision/uro_audiolab_panorama/panorama.jpg` created from 8 images (see `vision/uro_audiolab_panorama/raw`) recorded with a Canon EOS 70D. The software hugin was been used to combine the images. The hugin configuration file is also provided at `vision/uro-panorama/hugin_settings.pto` .

Fig. 6.1: 360° Panorama of the Audio Laboratory.

6.2 Stereo-Vision Capture from Adream Building, CNRS Toulouse


Published by members of the Two!Ears consortium

6.2.1 License

6.2.2 Description

This database contains views of the environment in the Adream-building, a robot-lab space consisting of a large hall that contains four small rooms without concrete windows, doors, and ceilings (Fig. 6.2). Impulse responses are available for this environment in *Two!Ears, CNRS Toulouse, Adream-building*.

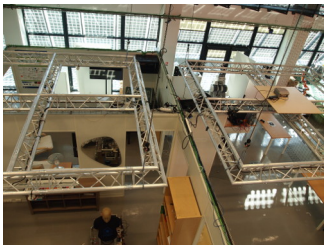


Fig. 6.2: Lab space in the Adream building, look from above.

The views are taken from the robot's perspective, using the calibrated stereovision system embedded on the robot (Fig. 6.3).

Views are available with the robot in all four positions (Fig. 6.4) for which impulse responses of where measured in the BRIR dataset *Two!Ears, CNRS Toulouse, Adream-building*. For positions the head orientation ranges from -78° to 78° in 2° steps. For each position, raw and rectified images are available from both left and right cameras.

6.2.3 Files

All raw images are available under for respective position (1-4) are available at:

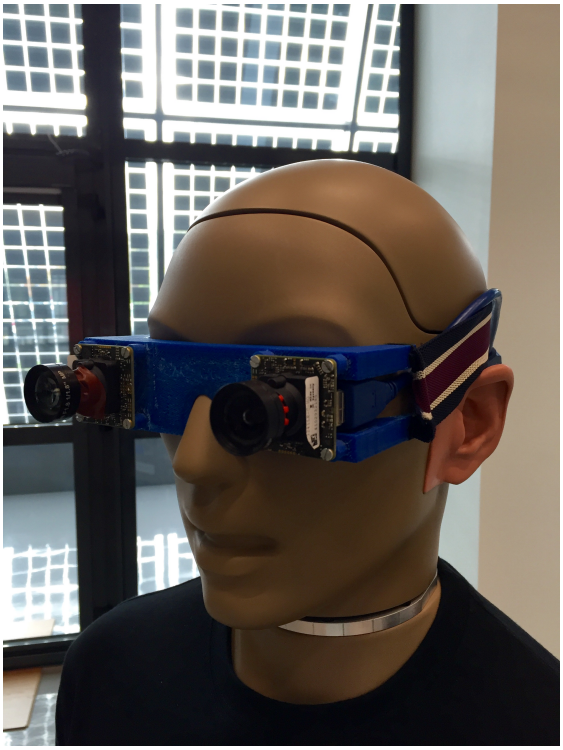


Fig. 6.3: KEMAR head with mounted stereo cameras.

```
vision/twoears_stereo_adream/raw/pos1/*
vision/twoears_stereo_adream/raw/pos2/*
vision/twoears_stereo_adream/raw/pos3/*
vision/twoears_stereo_adream/raw/pos4/*
```

Rectified images were obtained from the raw images with the stereo_image_proc tool from ROS (Robot Operating System). They are available under (same structure as raw images):

```
vision/twoears_stereo_adream/rect
```

The calibration data of the stereovision system includes text files with the estimated parameters. This data were computed and saved by the camera_calibration tool from ROS. It is available under:

```
vision/twoears_stereo_adream/calib
```

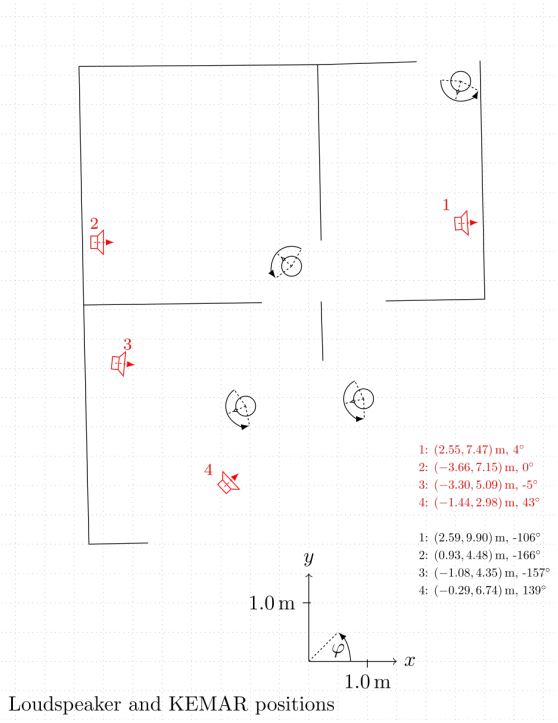


Fig. 6.4: Setup of the measurements. For details see the `PDF version` of this figure.

Bibliography

[Wierstorf2012] Wierstorf, H., Spors, S., Raake, A. (2012), "Perception and evaluation of sound fields," 59th Open Seminar on Acoustics, p. 263-68

[Wierstorf2014b] Wierstorf, H. (2014), "Perceptual Assessment of Sound Field Synthesis," PhD-thesis, TU Berlin

[KopcoEtAl2010] Kopco, N., Best, V., and Carlile, S. (2010) "Speech localization in a multitalker mixture," J. Acoust. Soc. Amer., vol. 127, no. 3, pp. 1450–1457.

[MaBrown2016] Ma, N., Brown, G. (2016) "SSpeech localisation in a multitalker mixture by humans and machines," In Proceedings of Interspeech 2016, San Francisco, CA.

[Wierstorf2014a] Wierstorf, H., Hohnerlein, C., Spors, S., Raake, A. (2014), "Coloration in wave field synthesis," 55th International Aes Conference, Paper 5-3

[Raake2014] Raake, A., Wierstorf, H., Blauert J. (2014), "A case for TWO!EARS in audio quality assessment," Forum Acusticum

[Hold2016a] Hold, C., Wierstorf, H., Raake, A. (2016), "The Difference Between Stereophony and Wave Field Synthesis in the Context of Popular Music," 140th AES Convention, Paper 9533

[Schultze2016] Schultze, A. (2016), "Der Sweet Spot in 5.0 Wiedergabesystemen in Abhängigkeit des Aufnahmeverfahrens und des visuellen Eindrucks des Zuhörers," bachelor thesis, Technische Universität Berlin

[Wittek2015] Wittek, H. (2015), "ORF Surround sound techniques, 2002," http://www.hauptmikrofon.de/stereo-3d/orf-surround-techniques, last access: 2016/10/21

[Hold2016b] Hold, C., Nagel, L., Wierstorf, H., Raake A. (2016), "Positioning of Musical Foreground Parts in Surrounding Sound Stages," AES International Conference on Audio for Virtual and Augmented Reality, Paper 7-2

[Wierstorf2011] Wierstorf, H., Geier, M., Raake, A., Spors, S. (2011) "A Free Database of Head-Related Impulse Response Measurements in the Horizontal Plane with Multiple Distances," 130th AES Convention, eBrief 6

[Bernschuetz2013] Bernschütz, B. (2013) "A Spherical Far Field HRIR/HRTF Compilation of the Neumann KU 100," German Annual Conference on Acoustics (DAGA)

[Bernschuetz2010] Bernschütz, B., Pörschmann, C., Spors, S., and Weinzierl, S. (2010) "Entwurf und Aufbau eines variablen sphärischen Mikrofonarrays für Forschungsanwendungen in Raumakustik und virtual Audio," German Annual Conference on Acoustics (DAGA)

[Gardner1994] Gardner, B., Martin, K. (1994) "HRTF measurements of a KEMAR dummy-head microphone," Massachusetts Institute of Technology 280

[Xie2013a] Xie, B. (2013), "Head-related transfer function and virtual auditory display," J Ross Publishing

[Xie2013b] Xie, B. et al. (2013), "Report on Research Projects on Head-Related Transfer Functions and Virtual Auditory Displays in China," Journal of the Audio Engineering Society (61) 5, pages 314-26

[Hummersone2011] Hummersone, C. (2011), "Binaural Room Impulse Response Measurements," pdf.

[Neidhardt2016] Neidhardt, A. (2016), "Perception of reververation captured in a real room, depending on position and direction", 22nd Int. Congress on Acoustics, Buenos Aires, Argentina