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

Deliverable 1.2

Intermediate database of audio-visual scenarios

WP 1 *

November 30, 2014

* 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:	D1.2
Document title:	Intermediate database of audio-visual scenarios
Version:	1.0

Delivery date:	30. November 2015
Actual publication date:	01. December 2015
Dissemination level:	Restricted
Nature:	Other

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

Contents

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 to generate 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 demands for a central database in order to provide access to this data among the member of the consortium and the public.

In the Deliverable D 1.1 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. The current infrastructure is documented in Section 2.

The data added to the database in the second year is documented in Section 3. While additional impulse response data sets from other research groups have been added, the database was also extended by three Binaural Room Impulse Response (BRIR) datasets measured by members of the consortium. Furthermore, new results from hearing experiments conducted in the last year including timbral perception and quality rating for different sound reproduction techniques have been added to the database.

2 Changes in the Infrastructure

As outlined in Deliverable 1.1, the Two!Ears database has been split into a public and a project-internal part. This is mainly due to restricted access of some data needed by the consortium members to develop the model. Both databases were managed by the version control software git[1]. Due to the noticeable growth of the project-internal database, 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 database although they might be only 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.
An extension to git called git-media[2] has now been used to overcome these problems. As shown in Fig. 2.1 the version control and storage of the data have now 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.

1 see `http://git-scm.com`
2 see `https://github.com/alebedev/git-media`

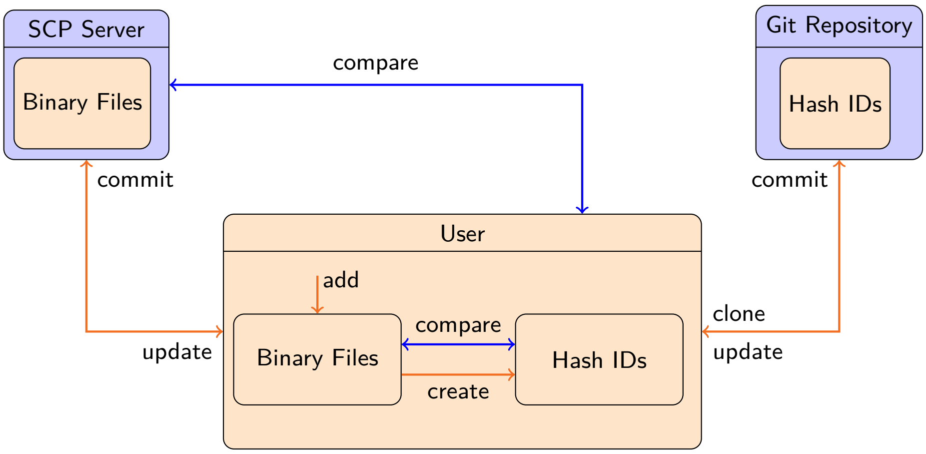


Figure 2.1: Structure of project-internal database

3 Content added to the Database

The Two!Ears database constitutes a central repository of data used for the synthesis and evaluation of the scenarios used by the consortium. The database is amended during the project. In the following an overview of content added to the database during the 2nd year is given. For the content of the first year, the reader is referred to D 1.1.

3.1 Impulse Responses

3.1.1 Head Related Impulse Responses

Database Entry #34

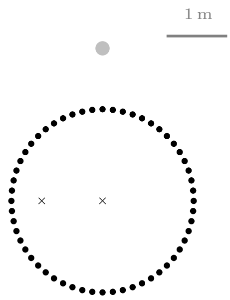
Title:	A high resolution head-related transfer function database including different orientations of head above the torso		
Path:	`impulse_responses/seacen_fabian_anechoic`		
Short:	SEACEN_FABIAN_ANECHOIC	**Reference:**	[1]
License:		**Access:**	Internal

Measurements were conducted in the anechoic chamber of the Carl von Ossietzky University, Oldenburg, Germany using the FABIAN [2] Head and Torso Simulator (HATS). During the measurement the head-above-torso orientation of the HATS has been changed to 11 different values. For each of these orientations the sound source was moved on a three-dimensional grid at a constant distance of approximately 1.7 metre. The elevation angle was varied from -64° to 90° with a resolution of 2°. The azimuth has varied to cover a full circle with a resolution depending on the current elevation (see table).

elevation	azimuth resolution
90°	360°
88°	45°
86° to 84°	18°
82° to 80°	10°
78°	9°
76° to 72°	6°
70° to 68°	5°
66° to 50°	3°
48° to -48°	2°
-50° to -64°	3°

Database Entry #35

Title:	Point source synthesized with Local Wave Field Synthesis and other Reproduction techniques using a 56-channel circular loudspeaker array		
Path:	`experiments/2015-10-05_localwfs_coloration/brs`		
Short:	BRS_LOCALWFS_PS_CIRCULAR...	**Reference:**	[1]
License:	CC BY-SA 4.0	**Access:**	Internal

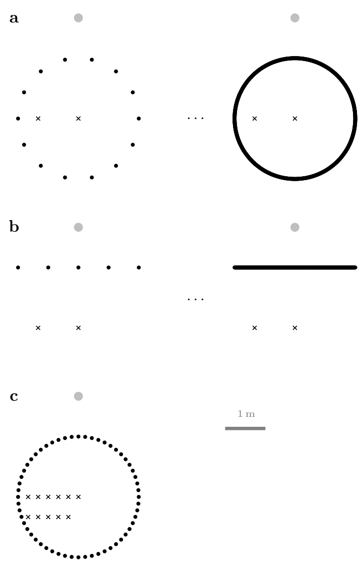


Data set of Binaural Room Synthesis (BRS) files used in a coloration experiment, containing impulse responses of a 56-channel loudspeaker array (black dots) driven by several reproduction techniques to synthesize a virtual point (grey dot). The reproduction methods are Near Field Compensated Higher Order Ambisonics (NFC-HOA), Wave Field Synthesis (WFS), and Local Wave Field Synthesis (LWFS) with various parametrisations. The impulse responses were computed separately for a centre and an off-centre position of the listener (crosses). The synthesis was done with the Head-Related Transfer Functions (HRTFs) from dataset #3, which mimic the impulse response from each loudspeaker to the listener's ears. The results of the listening experiment are provided with dataset #40.

1 Dataset is planned to be made public at the 140th Convention of the Audio Engineering Society, Paris, France

Database Entry #36

Title:	Point source synthesized with Wave Field Synthesis using different loudspeaker setups and listener positions.		
Path:	`experiments/2015-10-01_wfs_coloration/brs`		
Short:	BRS_WFS_PS_POSITIONS	**Reference:**	[1]
License:	CC BY-SA 4.0	**Access:**	Internal



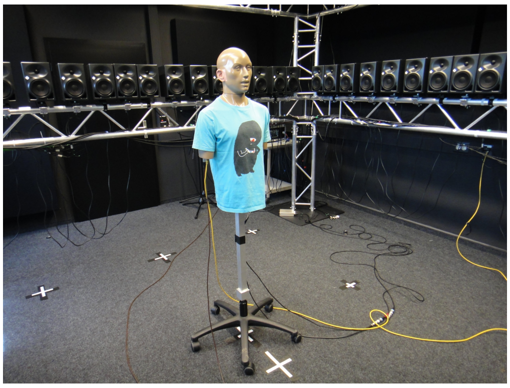
Data set of BRS files used in a coloration experiment, containing impulse responses of various reproduction setups driven by WFS to synthesize a virtual point source (see Fig., grey dot). In a) and b) a circular and a linear loudspeaker array (black dots) was used, respectively. The number of loudspeakers was varied in both cases. The impulse responses were computed separately for a centre and an off-centre position of the listener (crosses). In c) number of loudspeaker and the shape of the array was fixed, while the listener's position was varied (11 different positions). The synthesis was done with the HRTFs from dataset #3, which mimic the impulse response from each loudspeaker to the listener's ears. The results of the listening experiment are provided with dataset #41.

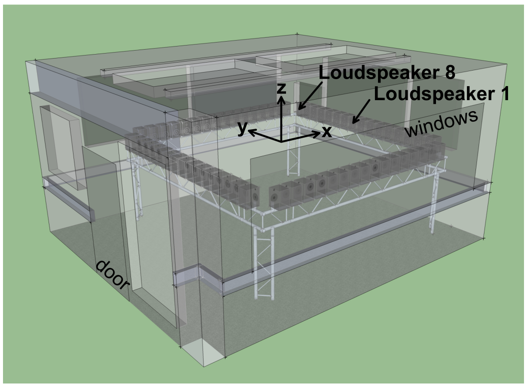
1 Not published yet

3.1.2 Binaural Room Impulse Responses

Database Entry #37

Title:	Single-channel RIRs and BRIRs of a 64-Channel Loudspeaker Array for different Room Configurations		
Path:	`impulse_responses/uro_kemar_audiolab`		
Short:	URO_KEMAR_AUDIOLAB	**Reference:**	[3]
License:	CC BY-NC-SA 4.0	**Access:**	Public

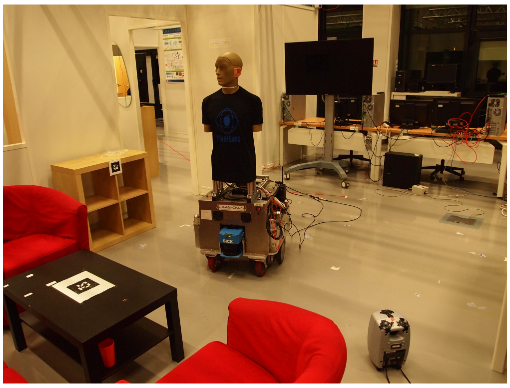




The database contains measured single-channel and binaural room impulse responses (RIRs and BRIRs) of a 64-channel loudspeaker array of square shape under varying room acoustical conditions. The measurements have been performed at the Audio Lab of the Institute of Communications Engineering, University 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. The data is provided in the Spatially Oriented Format for Acoustics (SOFA).

Database Entry #38

Title:	BRIRs from the KEMAR manikin in the Adream laboratory at LAAS CNRS, Toulouse, France		
Path:	`impulse_responses/twoears_kemar_adream`		
Short:	TWOEARS_KEMAR_ADREAM	**Reference:**	[1]
License:	CC BY-SA 4.0	**Access:**	Internal

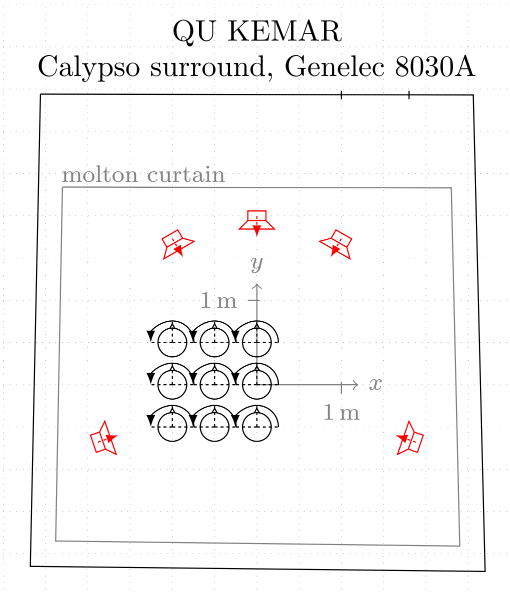


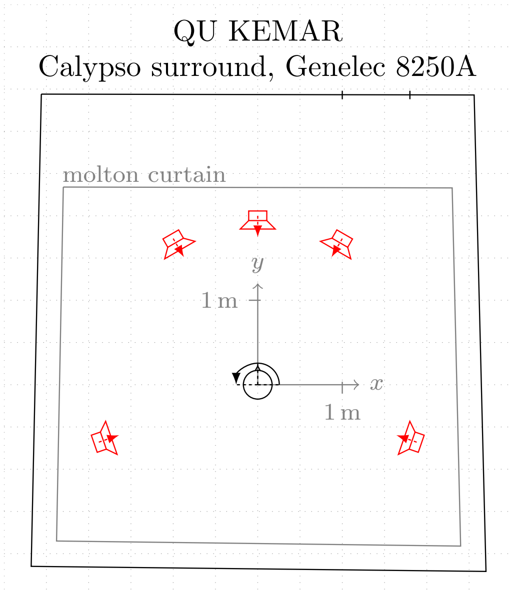
The BRIRs are measured using a KEMAR (type 45BB-4) HATS with the corresponding large pinnae (KB0091 + KB0090). The manikin was mounted on the movable robotics platform JIDO. Furthermore a servo motor was inserted into the HATS in order to turn its head. During the measurement the torso of the HATS had always the same orientation, only its head was rotated from -78° to 78° with a resolution of 2°. As sound sources four active two-way loudspeakers (Genelec 8020A) where placed at different positons inside the Adream Laboratory, Toulouse. The HATS was placed at four different positions, leading to a total number of 16 different source-position-listener-position combinations as illustrated above. Ear signals were recorded with G.R.A.S. 40AO 0.5 inch pressure microphones using a RME FIREFACE UC with an input gain of 11dB. All data was recorded with a sampling rate of 44.1kHz and stored as single precision floating point values.

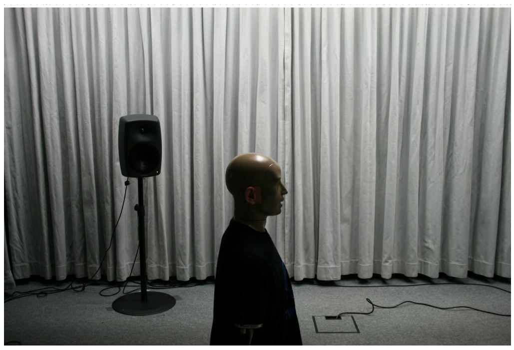
1 Dataset will be made public at the 140th Convention of the Audio Engineering Society, Paris, France

Database Entry #39

Title: BRIRs of 5.0 stereophonic setups with two different loudspeaker fabricates in studio room Calypso at TU Berlin
Path: `impulse_responses/qu_kemar_rooms/calypso_surround`
Short: QU_KEMAR_CALYPSO_SURROUND **Reference:** not published
License: CC BY-SA 4.0 **Access:** Internal







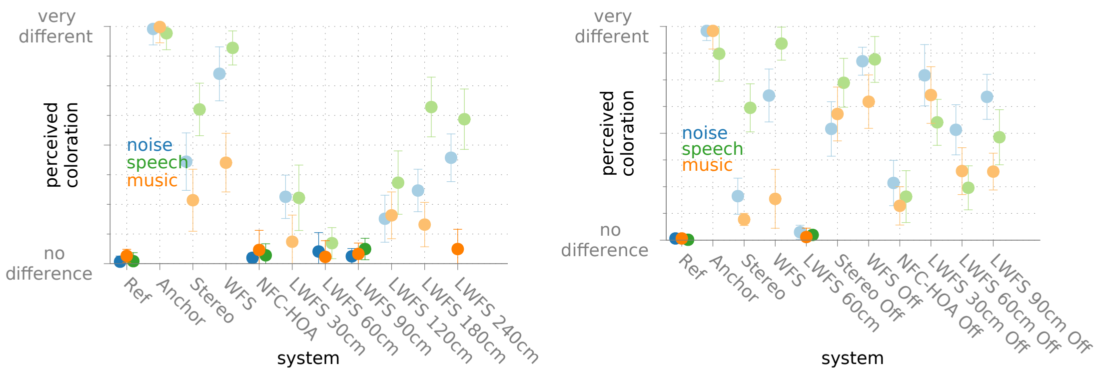
The BRIRs were measured using a KEMAR (type 45BA) with the corresponding large ears (type KB0065 and KB0066). 5 Loudspeakers were placed around the manikin to establish a 5.0 surround setup. The measurements were done with two different two-way loudspeaker fabricates, namely Genelec 8030A (left figures) and Genelec 8250A (right). For the former, the measurements were repeated at 9 different listening positions (top left). For each position of the manikin, its head was rotated horizontally from -90° to 90° with a resolution of 1°.

3.2 Perceptual Labels

3.2.1 Timbral perception for different spatial audio presentation techniques

Database Entry #40

Title:	Coloration for a point source synthesized with Local Wave Field Synthesis and other reproduction techniques using a 56-channel circular loudspeaker array		
Path:	`experiments/2015-10-05_localwfs_coloration`		
Short:	TIMBRAL_LOCALWFS_PS_CIRCULAR	**Reference:**	[1]
License:	CC BY-SA 4.0	**Access:**	Internal

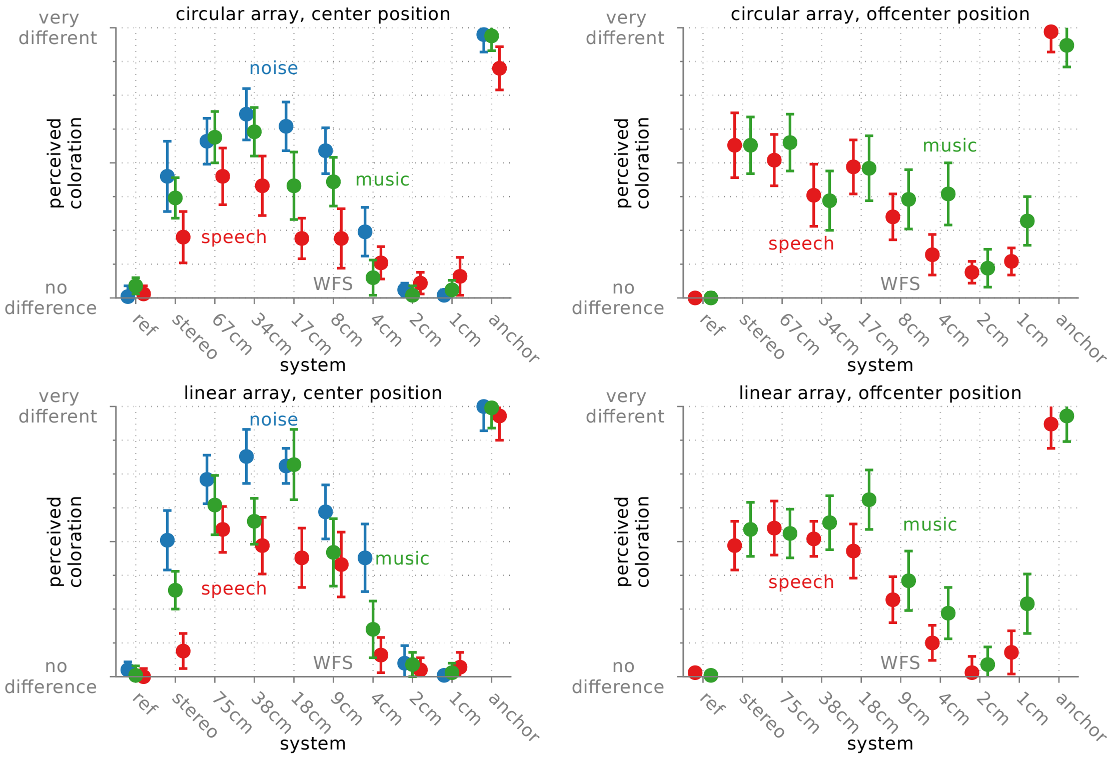


The experiment was run for several reproduction techniques in combination with two different listener positions. Three different source materials (speech, music, pink noise) were used. The dataset contains the individual responses from the subjects. The above plots shows a summary with median, 25% and 75% percentiles. A lower saturation of the colours indicates, that the colouration scores for the respective conditions differ significantly (significance level of 99%) from the reference. The corresponding BRS data set is provided as dataset #35.

1 Dataset is planned to be made public at the 140th Convention of the Audio Engineering Society, Paris, France

Database Entry #41

Title:	Coloration for a point source synthesized with Local Wave Field Synthesis and other Reproduction techniques using a 56-channel circular loudspeaker array		
Path:	`experiments/2015-10-01_wfs_coloration`		
Short:	TIMBRAL_WFS_PS_POSITIONS	**Reference:**	[1]
License:	CC BY-SA 4.0	**Access:**	Internal



Coloration in WFS for a central and an off-center listening position. The database contains the median over 16 listeners together with the confidence interval as shown in the plots. For the WFS conditions different circular and linear loudspeaker arrays were applied, where the used loudspeaker distances are marked at the tics of the x-axes. The corresponding BRS data set is provided as dataset #36.

1 parts of this have been presented in [4]

3.2.2 Quality ratings for different spatial audio presentation techniques

Database Entry #42

Title:	Preferred sound quality for a point source synthesized with different Wave Field Synthesis systems		
Path:	`experiments/2015-11-01_wfs_preferred_quality`		
Short:	QUALITY_WFS_PS	**Reference:**	not published
License:	CC BY-SA 4.0	**Access:**	Internal

For a central listening position in the different WFS systems used in the WFS coloration experiment #41 a complete paired comparison test was performed, asking the listeners after their preference regarding sound quality. The corresponding BRS data set is provided as dataset #36.

Database Entry #43

Title:	Multi-dimensional scaling for a point source synthesized with different Wave Field Synthesis systems		
Path:	`experiments/2015-11-15_wfs_mds`		
Short:	MDS_WFS_PS	**Reference:**	not published
License:	CC BY-SA 4.0	**Access:**	Internal

For the same central listening position and the different WFS systems used in experiment #42, a multi-dimensional scaling experiment was performed. In this experiment listeners were presented with all possible pairs of stimuli and had to judge their perceptual difference on a continuous scale. The corresponding BRS data set is provided as dataset #36.

Acronyms

BRIR Binaural Room Impulse Response 3, 10–12

BRS Binaural Room Synthesis 8, 9, 13–15

CC BY-NC-SA 4.0 Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International, see `http://creativecommons.org/licenses/by-nc-sa/4.0/legalcode` 10

CC BY-SA 4.0 Creative Commons Attribution-ShareAlike 4.0 International, see `http://creativecommons.org/licenses/by-sa/4.0/legalcode` 8, 9, 11–15

HATS Head and Torso Simulator 7, 11

HRTF Head-Related Transfer Function 8, 9

KEMAR Knowles Electronics Manikin for Acoustic Research 11, 12

LWFS Local Wave Field Synthesis 8

NFC-HOA Near Field Compensated Higher Order Ambisonics 8

RIR Room Impulse Response 10

SCP Secure Copy 5

SOFA Spatially Oriented Format for Acoustics 10

WFS Wave Field Synthesis 8, 9

Bibliography

[1] F. Brinkmann, A. Lindau, S. Weinzierl, G. Geissler, and S. van de Par, “A high resolution head-related transfer function database including different orientations of head above the torso,” in *Proceedings of the AIA-DAGA 2013 Conference on Acoustics*, 2013. (Cited on page 7)

[2] A. Lindau and S. Weinzierl, “Fabian-schnelle erfassung binauraler raumimpulsantworten in mehreren freiheitsgraden,” *Fortschritte der Akustik*, vol. 33, no. 2, p. 633, 2007. (Cited on page 7)

[3] V. Erbes, M. Geier, S. Weinzierl, and S. Spors, “Database of single-channel and binaural room impulse responses of a 64-channel loudspeaker array,” in *Audio Engineering Society Convention 138*, May 2015. [Online]. Available: http://www.aes.org/e-lib/browse.cfm?elib=17624 (Cited on page 10)

[4] J. Ahrens and H. Wierstorf, “Recent advancements in massive multi-channel auralization,” *The Journal of the Acoustical Society of America*, vol. 138, no. 3, 2015. (Cited on page 14)