Audio Engineering Society
Convention e-Brief 330
Presented at the 142nd Convention
2017 May 20–23, Berlin, Germany

The TWO!EARS Database

Fiete Winter[1], Hagen Wierstorf[2], Alexander Raake[2], and Sascha Spors[1]

[1]*Institute of Communications Engineering, University of Rostock, Rostock, D-18119, Germany*
[2]*Audiovisual Technology Group, Technische Universität Ilmenau, Ilmenau, D-98693, Germany*

Correspondence should be addressed to Fiete Winter (`fiete.winter@uni-rostock.de`)

ABSTRACT

TWO!EARS was an EU-funded project for binaural auditory modelling with ten international partners involved. Its main goal was to provide a computational framework for the modelling of active exploratory listening that assigns meaning to auditory scenes. As one outcome of the project, a database including data acquired by the involved partners as well as third-party measurements has been published. Among others, a large collection of Head Related Impulse Responses and Binaural Room Impulse Responses is part of the database. Further, results from psychoacoustic experiments conducted within TWO!EARS to validate the developed auditory model were added. For the usage of the database together with the TWO!EARS model, a software interface was developed to download the data from the database on demand.

1 Introduction

One of the outstanding capabilities of the human auditory system is to recover information on single auditory objects out of a mixture of sounds [1]. TWO!EARS was an EU FET-OPEN project funded until December 2016 and aimed at developing a computational model that mimics this behaviour. Listeners are regarded as multi-modal agents that develop their concept of the world by active, exploratory sensing. In the course of this process, they interpret percepts, applying existing and collecting new knowledge, and concepts accordingly. Consequently, the TWO!EARS model involves bottom-up (signal-driven) as well as top-down (hypothesis-driven) processes. As the two major application areas of the model, Dynamic Auditory Scene Analysis (DASA) and Quality of Experience (QoE) were approached during the project. Former subsumes the formation of auditory objects by humans' based on the ear signals. Latter implies quality assessment of spatial sound reproduction techniques such as traditional stereo and Wave Field Synthesis (WFS) [2].

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 synthesis of ear signals and eye images is an important basis for the development and evaluation of the TWO!EARS 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 project members and the scientific community.

This engineering brief provides information about the TWO!EARS database. It is organised as follows: in

Sec. 2, an overview about the content of the database is given. Information about the accessibility of the database are described in Sec. 3, followed by a summary in Sec. 4.

2 Content of the Database

In this section, the content of the Two!Ears database is described on a macroscopic level. For a more detailed, extensive description of particular datasets, the reader is referred to the online documentation[1] of the database. Most of the datasets acquired by the Two!Ears consortium have also been uploaded to the online platform Zenodo[2] assigning citable Digital Object Identifiers (DOIs) to the data. If applicable, license information concerning the (re-)publication and modification of the data is given.

2.1 Impulse Responses

For the auralisation and rendering of acoustic scenes using binaural synthesis, 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 Head-Related Impulse Responses (HRIRs). The measurement of this impulse responses is usually done with a Head and Torso Simulator (HATS), which models the human outer ears. While HRIRs imply free-field conditions and an anechoic environment, Binaural Room Impulse Responses (BRIRs) include reverberation caused by the obstacles and walls. The Two!Ears database provides a large collection of HRIRs and BRIRs. Some of those were measured by members of the Two!Ears project: a BRIR dataset was captured in an apartment-like environment [3] which may be used in the DASA application. For the QoE task, a BRIR dataset of a 64-channel loudspeaker array measured under different acoustic conditions [4] can be utilised. All datasets are available in the Spatially Oriented Format for Acoustics (SOFA), which is the standard for spatial acoustic data of the Audio Engineering Society [5]. This format is also supported by the binaural simulator contained within the Two!Ears model [6].

2.2 Results from Listening Tests

Several listening tests where conducted during the project in order to provide the model with perceptual

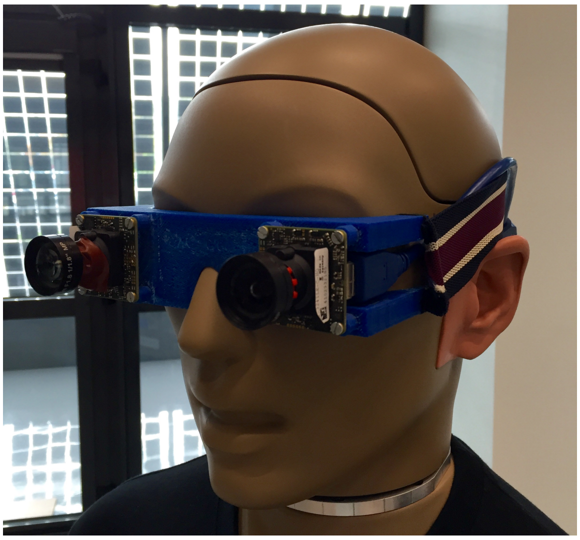

Fig. 1: Knowles Electronics Manikin for Acoustic Research (KEMAR) [7] Head and Torso Simulator used for the BRIR measurements in [3] with mounted stereo cameras.

labels and train or validate its output against those. The database includes the results from localisation experiments [8, 9, 10] using HRIR-based binaural simulation of various acoustic scenarios. Further, colouration ratings, e.g. from [11], are included: here, different spatial sound reproduction techniques were rated in terms of their perceived colouration compared to a reference using a modified Multi-Stimulus with Hidden Anchor and Reference (MUSHRA) [12] test paradigm. As the third major part, preference ratings [15, 14, 13] to assess the quality of different spatial sound reproduction techniques are comprised. All datasets include the used dry stimuli and, if necessary, impulse responses for the binaural simulation.

2.3 Acoustic & Visual Stimuli

In HRIR/BRIR-based binaural synthesis, dry source signals are of great importance, as they are filtered by the impulse responses corresponding to the sound source positions. The database contains a subset of a multitalker audiovisual sentence corpus [16], noise stimuli, and instrumental stimuli. Further, a dataset [17] of binaural recordings from acoustic environments, e.g. office, restaurant, or tube station, was included. The huge variety of different stimuli together with the

[1] http://docs.twoears.eu/en/latest/database/

[2] https://zenodo.org/

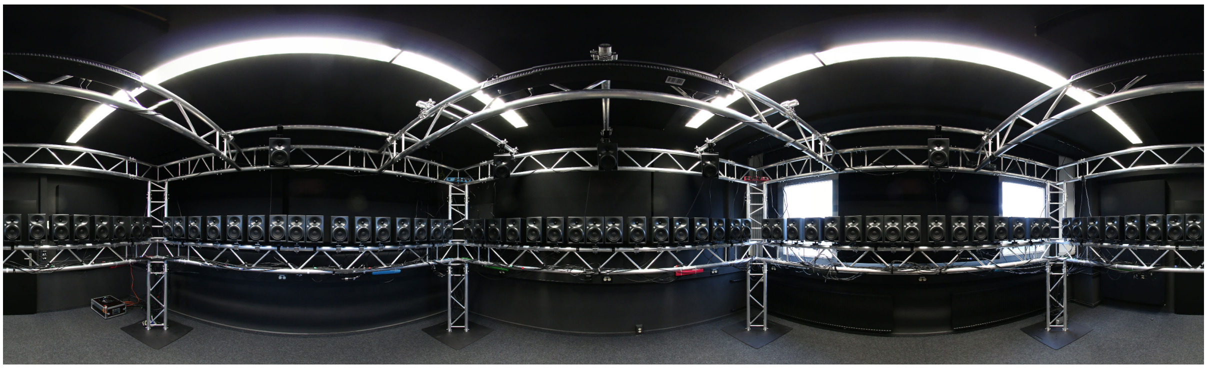

Fig. 2: Panorama of the audio laboratory at the Institute of Communications Engineering, University of Rostock

large number of impulse responses allows for the synthesis of many acoustic scenarios, which are necessary to train and evaluate perceptual models.

In order to incorporate multi-modality into perceptual modelling, a small selection of visual stimuli is part of the database. Corresponding to the BRIRs [3] mentioned in Sec. 2.1, stereoscopic images are available for all positions for which BRIRs were measured. The images where recorded with a stereo-camera pair mounted on the HATS, cf. Fig. 1. For each position, raw and rectified images are available from both left and right cameras. Hence is it possible to combine the acoustic simulation with stereoscopic views. Furthermore, a cylindrical indoor panorama of an audio laboratory, cf. Fig. 2, corresponding to one measured position for the BRIR dataset in [4], was added.

3 Accessing the Database

The infrastructure of the database is grounded on the project management web application Redmine[3]. Using a suitable HTML client, the user can browse through the data[4] and download particular files. All files are version controlled using Apache Subversion (SVN). The repository can be checked out using the URL[5]. SVN also supports partial check outs of subdirectories by appending their path relative to the root of repository to the URL. The paths for specific datasets can found in respective entry in the online documentation.

A MATLAB interface to download the data on demand is shipped as part of the software package for the Two!Ears model [6]. A needed file can be downloaded on demand using `fname = db.getFile('relative/path/to/file')`, where `fname` contains the name of the locally stored file after the download. The relative path of the file is the same as for the SVN repository and is also listed in the respective documentation entry. In order to prevent unnecessary downloads, a flexible file inclusion mechanism takes different possible locations (in order of appearance) of the needed file into account: (1) Search the file relatively to the current working directory which is determined via the built-in function `pwd()`. (2) Search the file relatively to the root of the file system. (3) Search inside a local copy of the whole database repository, which has been downloaded via SVN. The location of the local copy can be defined using `db.path('absolute/path/to/local/copy')`. (4) Search inside a temporary directory, which caches single files downloaded from the database before, in order to prevent repeating downloads. The location of the temporary directory can be set via `db.tmp('absolute/path/to/tmp/dir')`. If desired, the temporary directory can be cleared using `db.clearTmp()`.(5) The file is downloaded from the remote repository, whose URL can be redefined by `db.url('http://url.database.eu')`. The downloaded file will be stored inside the temporary directory. It is furthermore possible to request entire directories from the database with `db.getDir('relative/dir')`. To get an a-priori overview about the content of an directory, `db.listDir('relative/dir')` can be called.

4 Summary

With regard to the concept to Open Science including Open Data, the Two!Ears database is a significant

[3] http://www.redmine.org/

[4] https://dev.qu.tu-berlin.de/projects/twoears-getdata

[5] https://dev.qu.tu-berlin.de/svn/twoears-getdata

contribution to the acoustic research community. Clear licensing information and extensive documentation for each dataset foster their usage in future scientific work, e.g. on spatial hearing or machine listening. The usage of DOIs ease the referencing within publications. Future work will address the process of extending the database and the auditory model beyond TWO!EARS to make these public domain and respective contributions attributable to the authors, not the project. In the view of the authors, here a more neutral labelling will motivate other researchers to contribute to the database with their own work.

5 Acknowledgements

This research has been supported by EU FET grant TWO!EARS, ICT-618075.

References

[1] Cherry, E. C., "Some experiments on the recognition of speech, with one and with 2 ears," *J. Acoust. Soc. Am.*, 25, pp. 975–979, 1953.

[2] Berkhout, A. J., "A Holographic Approach to Acoustic Control," *J. Aud. Eng. Soc.*, 36(12), pp. 977–995, 1988.

[3] Winter, F., Wierstorf, H., Podlubne, A., Forgue, T., Manhès, J., Herrb, M., Spors, S., Raake, A., and Danès, P., "Database of binaural room impulse responses of an apartment-like environment," in *Proc. of 140th Aud. Eng. Soc. Conv.*, Paris, France, 2016.

[4] Erbes, V., Geier, M., Weinzierl, S., and Spors, S., "Database of Single-Channel and Binaural Room Impulse Responses of a 64-Channel Loudspeaker Array," in *Proc. of 138th Aud. Eng. Soc. Conv.*, 2015.

[5] Audio Engineering Society, Inc., "AES69-2015 - AES standard for file exchange - Spatial acoustic data file format," 2015.

[6] Two!Ears Team, "Two!Ears Auditory Model 1.4," 2017, doi:10.5281/zenodo.238761.

[7] G.R.A.S. Sound & Vibration, "Instruction Manual - G.R.A.S. 45BB KEMAR Head and Torso / 45BC KEMAR Head and Torso with Mouth Simulator," 2016.

[8] Wierstorf, H., Spors, S., and Raake, A., "Perception and evaluation of sound fields," in *59th Open Seminar on Acoustics*, 2012.

[9] Wierstorf, H., *Perceptual Assessment of Sound Field Synthesis*, Ph.D. thesis, Technische Universität Berlin, 2014.

[10] Ma, N. and Brown, G. J., "Speech Localisation in a Multitalker Mixture by Humans and Machines," in *Interspeech 2016*, pp. 3359–3363, 2016.

[11] Wierstorf, H., Hohnerlein, C., Spors, S., and Raake, A., "Coloration in Wave Field Synthesis," in *Proc. of 55th Intl. Aud. Eng. Soc. Conf. on Spatial Audio*, Helsinki, Finland, 2014.

[12] International Telecommunication Union Radiocommunication Assembly, "ITU-R BS.1534-3 - Method for the subjective assessment of intermediate quality level of audio systems," 2015.

[13] Hold, C., Nagel, L., Wierstorf, H., and Raake, A., "Positioning of Musical Foreground Parts in Surrounding Sound Stages," in *Proc. of Int. Aud. Eng. Soc. Conf. on Audio for Virtual and Augmented Reality*, 2016.

[14] Hold, C., Wierstorf, H., and Raake, A., "The difference between stereophony and wave field synthesis in the context of popular music," in *Proc. of 140th Aud. Eng. Soc. Conv.*, Audio Engineering Society, 2016.

[15] Raake, A., Wierstorf, H., and Blauert, J., "A case for TWO!EARS in audio quality assessment," in *Forum Acusticum*, 2014.

[16] Cooke, M., Barker, J., Cunningham, S., and Shao, X., "An audio-visual corpus for speech perception and automatic speech recognition," *J. Acoust. Soc. Am.*, 120(5), pp. 2421–2424, 2006.

[17] Giannoulis, D., Benetos, E., Stowell, D., Rossignol, M., Lagrange, M., and Plumbley, M. D., "Detection and classification of acoustic scenes and events: An IEEE AASP challenge," in *2013 IEEE Workshop on Applications of Signal Processing to Audio and Acoustics*, 2013, doi:10.1109/WASPAA.2013.6701819.