Convention E-Brief 330

Fiete Winter[1], Hagen Wierstorf[2], Alexander Raake[2], Sascha Spors[1]
[1]Institute of Communications Engineering, Universität Rostock
[2]Audiovisual Technology Group, Technische Universität Ilmenau
fiete.winter@uni-rostock.de

The TWO!EARS Database

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. It 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.

Content of the Database

- detailed documentation of the data available at

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

- if applicable, license information is given
- citable Digital Object Identifiers (DOIs) are assigned to most of the datasets acquired by the TWO!EARS consortium

Impulse Responses

All impulse response are stored in the Spatially Oriented Format for Acoustics (SOFA) [3]

Head-Related Impulse Responses (HRIRs)

- Anechoic HRIRs from the Knowles Electronics Manikin for Acoustic Research (KEMAR) with different distances [4]
- Spherical far-field HRIR compilation of the Neumann KU100 [5]
- MIT HRIR measurements of a KEMAR dummy head [6]
- Near-field HRIRs from SCUT database of the KEMAR [7]

Binaural Room Impulse Responses (BRIRs)

- TWO!EARS, CNRS Toulouse, Adream-building [8], see Figure 2
- TU Berlin, different rooms (Auditorium 3, Spirit, and Calypso)
- University of Rostock, Room Impulse Responses (RIRs) and BRIRs of a 64-channel loudspeaker array for different room configurations [9], see Figure 3
- Salford-BBC, 12-channel loudspeaker studio
- University of Surrey, four different rooms
- TU Ilmenau, conference room [10]

Results from Listening Tests

- localisation experiments [11, 12, 13] using HRIR-based binaural simulation of various acoustic scenarios, e.g. Figure 1
- colouration ratings of different spatial sound reproduction techniques [14] using a Multi-Stimulus with Hidden Anchor and Reference (MUSHRA) [15] test paradigm
- preference ratings [16, 17, 18] to assess the quality of different spatial sound reproduction techniques

Acoustic & Visual Stimuli

- subset of a multitalker audiovisual sentence corpus [19]
- binaural recordings from acoustic environments [20]
- stereo-vision capture from Adream Building, LAAS-CNRS Toulouse corresponding to the BRIR measurements [8], see Figure 2c
- cylindrical indoor panorama of an audio laboratory corresponding to one measured position for the BRIR dataset in [9], see Figure 3

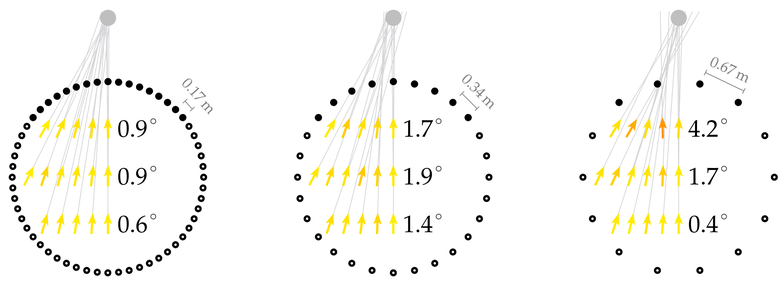
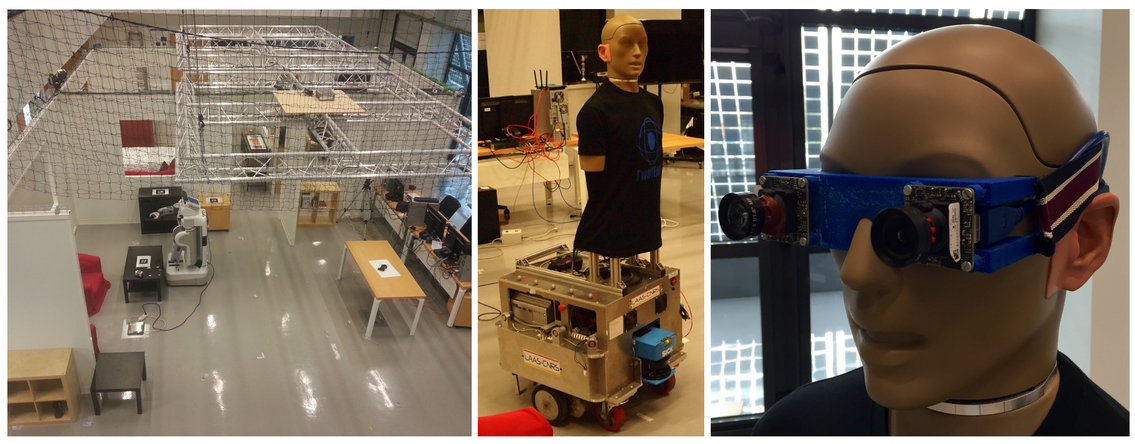


Figure 1: Average localisation results for a circular array (black circles) of 56, 28, and 14 loudspeakers (left to right) driven with WFS to synthesise a virtual point source (grey dot). For each listening position, an arrow points towards the perceived sound source direction. Details can be found in [12].



(a) (b) (c)

Figure 2: (a) ADREAM Laboratory, LAAS-CNRS, Toulouse. (b) Jido mobile robot with the KEMAR mounted on its top used for BRIR measurements [8]. (c) KEMAR Head and Torso Simulator with mounted stereo cameras.

Accessing the Database ...

... via web browser

- the user can browse through the data and download particular files at

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

... via Apache Subversion (SVN)

- all files are version controlled using SVN
- the repository can be checked out using a suitable SVN client together with

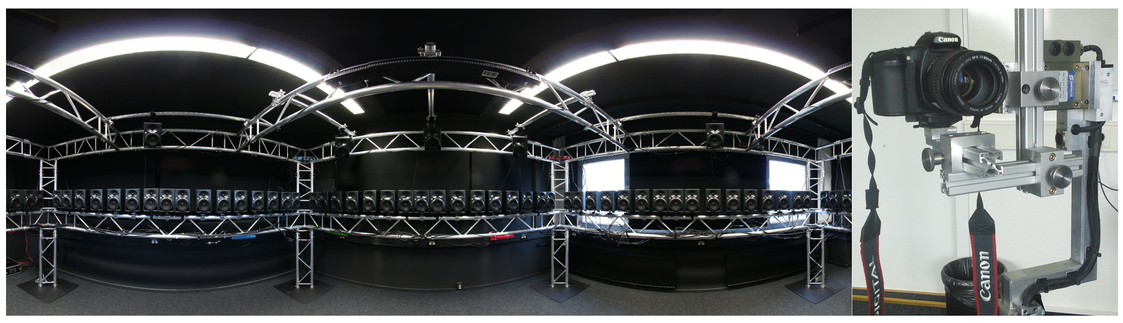
  https://dev.qu.tu-berlin.de/svn/twoears-getdata

- SVN also supports partial check outs of subdirectories
- the paths for specific datasets can be found in the respective entry of the online documentation

... via MATLAB

- an interface to download the data on demand is shipped as part of the software package for the TWO!EARS model [21] which is available at

  http://twoears.eu



(a) (b)

Figure 3: (a) Panorama of the audio laboratory at the Institute of Communications Engineering, University of Rostock, where the BRIR measurements of [9] were conducted. (b) Apparatus to automatically capture an indoor panorama.

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] Audio Engineering Society, Inc., "AES69-2015 - AES standard for file exchange - Spatial acoustic data file format," 2015.
[4] Wierstorf, H., Geier, M., and Spors, S., "A Free Database of Head Related Impulse Response Measurements in the Horizontal Plane with Multiple Distances," in *Proc. of 130th Aud. Eng. Soc. Conv.*, London, UK, 2011.
[5] Bernschütz, B., "A Spherical Far Field {HRIR}/{HRTF} Compilation of the Neumann {KU} 100," in *German Annual Conference on Acoustics (DAGA)*, Meran, Italy, 2013.
[6] Gardner, B. and Martin, K., "HRTF Measurements of a KEMAR Dummy-Head Microphone," Technical Report 280, MIT Media Lab, 1994.
[7] Xie, B., *Head-Related Transfer Function and Virtual Auditory Display*, J Ross, 2013.
[8] 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.
[9] 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.
[10] Neidhardt, A., "Perception of reververation captured in a real room, depending on position and direction," in *Proc. of 22nd Int. Congress on Acoustics*, Buenos Aires, Argentina, 2016.
[11] Wierstorf, H., Spors, S., and Raake, A., "Perception and evaluation of sound fields," in *59th Open Seminar on Acoustics*, 2012.
[12] Wierstorf, H., *Perceptual Assessment of Sound Field Synthesis*, Ph.D. thesis, Technische Universität Berlin, 2014.
[13] Ma, N. and Brown, G. J., "Speech Localisation in a Multitalker Mixture by Humans and Machines," in *Interspeech 2016*, pp. 3359–3363, 2016.
[14] 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.
[15] International Telecommunication Union Radiocommunication Assembly, "ITU-R BS.1534-3 - Method for the subjective assessment of intermediate quality level of audio systems," 2015.
[16] Raake, A., Wierstorf, H., and Blauert, J., "A case for TWO!EARS in audio quality assessment," in *Forum Acusticum*, 2014.
[17] 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.
[18] 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.
[19] 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.
[20] 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.
[21] Two!Ears Team, "Two!Ears Auditory Model 1.4," 2017, doi:10.5281/zenodo.238761.