Evaluating the auralization of a small room in a virtual sound environment using objective room acoustic measures

Ahrens, Axel; Marschall, Marton; Dau, Torsten

Publication date: 2016

Document Version
Publisher's PDF, also known as Version of record

Link back to DTU Orbit

Citation (APA):
Evaluating the auralization of a small room in a virtual sound environment using objective room acoustic measures

Axel Ahrens, Marton Marschall, Torsten Dau
Hearing Systems group, Technical University of Denmark, Lyngby, Denmark
aahr@elektro.dtu.dk

Introduction
To study human auditory perception in realistic environments, loudspeaker-based reproduction techniques have recently become state-of-the-art. To evaluate the accuracy of a simulation-based room auralization of a small room, objective measures were evaluated in particular: early-decay time (EDT), reverberation time (T20, T30), clarity (CT), C50, C60), interaural cross-correlation (ACC), speech transmission index (STI), direct-to-reverberant ratio (DRR), impulse responses (IRs) were measured in an IEC listening room. The room was then modeled in the room acoustic software ODEON, and the sound objective measures were evaluated for synthesized versions of the playback room. The auralizations were realized using higher-order ambisonics (HOA), mixed-order ambisonics (MOA), and a nearest loudspeaker method (NL) and reproduced in a virtual sound environment.

Room Acoustic Measures

Figure 1: Reverberation time (T20; measured/standard deviation) over octave bands, measured at 7 source and 4 receiver positions. The blue and red curves indicate the ODEON model and the reference room, respectively. The remaining curves are auralized versions of the room.

Figure 2: Clarity (C50; measured/standard deviation) over octave bands, measured at 7 source and 4 receiver positions. The blue and red curves indicate the ODEON model and the reference room, respectively. The remaining curves are auralized versions of the room.

Figure 3: Root-mean-square error (RMSE) of the clarity decay measures relative to the ODEON model. The dashed line indicates the perceptual just noticeable difference (JND; 0.03; scale factor 10:1).

Binaural Measures

Figure 4: Root-mean-square error (RMSE) of the IACC measures relative to the reference. The early IACC is calculated over the first 2 ms of the impulse response. The dashed line indicates the perceptual just noticeable difference (JND; 0.03; scale factor 10:1).

Figure 5: Root-mean-square error (RMSE) of the energy decay measures relative to the reference. The dashed line indicates the perceptual just noticeable difference (JND; 0.03; scale factor 10:1).

Binaural Direct-to-Reverberant Ratio

Figure 6: Binaural direct-to-reverberant ratio (DRR) for the depicted source/receiver combinations, recorded with the left ear at a head-and-earimulator as a function of playback position. The blue and red curves indicate the ODEON model and the reference room, respectively. The remaining curves are auralized versions of the room.

Figure 7: Distribution of perceptual DRR (y-axis) and IACC (y-axis) calculated for a source 30° right of the IACC (new pictogram). Inaudible differences were analyzed in 20 ms windows with 10% overlap over 1 s long playbacks. Inaudible differences were calculated for the depicted octave bands.

Speech Intelligibility and STI

Figure 8: Speech transmission index (STI) parameters (left): measured and estimated at 7 source and 4 receiver positions (blush markers). The dashed line indicates the perceptual just noticeable difference (JND; 0.03; scale factor 10:1).

Conclusions

- Long-term, averaged measures are reproduced in the range of ~1 JND (T20/30, C50/80, STI, IACC).
- Short-term features of the impulse response are more difficult to capture leading to higher errors in e.g. EDT and CT.
- Similar performances were obtained across reproduction techniques.
- Auralization errors (auralization vs. model) are in the range of modelling errors (model vs reference).
- Dynamic binaural cues appear to be well captured.

Perceptual differences (e.g. speech intelligibility) occur, but not reflected in shown objective measures.

Further investigations needed to link perceptual differences to objective measures.

Literature


Binaural Direct-to-Reverberant Ratio

Figure 8: Binaural direct-to-reverberant ratio (DRR) for the depicted source/receiver combinations, recorded with the left ear at a head-and-earimulator as a function of playback position. The blue and red curves indicate the ODEON model and the reference room, respectively. The remaining curves are auralized versions of the room.

Acknowledgements

The authors would like to thank the developers of the ITP toolbox from RWTH Aachen as well as the maintenance of the Two!Ears Framework. Additionally, thanks to Oticon Research Centre, Eindhoven, for providing the speech material of the Dantale II. Thanks to the developers of the Two!Ears Framework.

This project was jointly funded by the Danish Council for Excellence in Hearing and Speech Sciences (DCHS).

Method

Reference Room
Room Acoustic Model
Virtual Sound Environment

Reproduction techniques
- Nearest loudspeaker (NL), Favrot&Buchholz, 2010
- Higher-order ambisonics (HOA, 5th order)
- Mixed-order ambisonics (MOA, 7th/5th order; Daniel, 2003)

Modeling
- ODEON v13.04 (Favrot&Buchholz, 1993) model of IEC listening room (7.5°, 7.5°, 7.8 m)
- Material properties optimized using ODEON’s genetic material optimizer (Christensen et al., 2014)

IR recording
- 7 source positions (Dynaudio BM8)
- 4 receiver positions (B&K 4192 and B&K HATS Type 4100)

- Processing and analysis using ITA-toolbox and Two!Ears framework

- Computational Toolbox from RWTH Aachen as well as the maintenance of the Two!Ears Framework.

- Estimating absorption of materials to match room model against existing room using a genetic algorithm. Forum Acusticum, Pola

- A Hybrid Computer Model for Room Acoustic Modelling. 4th Western Pacific Regional Acoustics Conference, Brisbane. Proc

- Speech reception thresholds (SRT) as target loudness for each condition. The target speech was presented from 0° and 2° masker angles.

- Root mean square error (RMSE) of the IACC measures relative to the reference. The early IACC is calculated over the first 2 ms of the impulse response. The dashed line indicates the perceptual just noticeable difference (JND; 0.03; scale factor 10:1).

- Speech transmission index (STI; mean±standard deviation)

- Conceptual Tools, A Comparative Study, A CHeSS study by the Odeon research team (2015).

- Speech intelligibility was assessed using the Danish Dantale II database.

- Speech intelligibility and STI

- Speech transmission index (STI) parameters (left): measured and estimated at 7 source and 4 receiver positions (blush markers). The dashed line indicates the perceptual just noticeable difference (JND; 0.03; scale factor 10:1).

- The grey bar shows the perceptual just noticeable difference (JND; 0.03; scale factor 10:1).