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, C80)
- interaural cross-correlation (IACC)
- speech transmission index (STI)
- direct-to-reverberant ratio (D2R)

Impulse responses (IRs) were measured in an IEC-listening room. The room was then modeled in the room-acoustic software ODEON, and the same objective measures were evaluated for simulated versions of the playback room. The auralizations were realized using higher-order ambitions (HOA), mixed-order ambitions (MOA), and a neural-loudspeaker method (NL) and reproduced in a virtual sound environment.

Method

Reference Room
Room Acoustic Model
Virtual Sound Environment

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

Modeling
- ODEON version 13.04 (Phonomega, 1995): model of IEC listening room (7.5%/75%/2.8m)
- Material properties optimized using ODEON's genetic material optimizer (Christensen et al., 2014)
IR recording
- 7 source positions (Dynaudio BMI)
- 4 receiver positions (B&K 4192 and B&K HATS Type 4100)
- Processing and analysis using ITA-toolbox and TwoEars framework

Room Acoustic Measures

Binaural Measures

Binaural Direct-to-Reverberant Ratio

Speech Intelligibility and STI

Figure 1: Reverberation time (T20) measured (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 averaged versions of the room.

Figure 2: Clarity (C80 measured 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 averaged versions of the room.

Figure 3: Root mean square error (RMSE) of the clarity measures relative to the ODEON model. The dashed line indicates the perceptual just noticeable difference for C80 in dB (Avarin et al., 2011).

Figure 4: Root mean square error (RMSE) of the reverberation time measures relative to the ODEON model. The dashed line indicates the perceptual just noticeable difference for reverberation time (0.06; Avarin et al., 2011).

Acknowledgements
The authors would like to thank the developers of the ITA-toolbox from RWTH Aachen as well as all the developers of the TwoEars Framework, Binaural Hall, and the Acoustics Research Centre Erlangen, for providing the speech material of the German 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 modeling 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

- Avarin et al. (2011): Speech transmission index (STI) characterization based on 7 source and 4 receiver positions (English). The authors indicate that the STI between these two receivers and a larger speaker location is depicted in Fig. 10.