STQ Workshop

Recreating complex soundscapes for audio quality evaluation

Presented by Jan Banas (Intel Corporation)

21/11/2022

Agenda

- Motivation
- State of the art
- Recording
- Processing
- Playback
- Conclusion

Motivation

- With speech-enabled devices design, there is a need for repeatable, realistic and spatially accurate background noise reproduction
- Noise suppression algorithms need to be evaluated in difficult, life-like scenarios
- The quality of background noise transmission contributes to the perceived overall quality of speech

State of the art

- ETSI ES 202 396-1
  - Solution designed for playback of binaural recordings through an array of loudspeakers
  - Calibrated using a HATS, commonly used for non-HATS devices, e.g., smart speakers or laptops
  - The simpler loudspeakers layout is a great advantage and has been adopted by industry

State of the art

- ETSI TS 103 224
  - Solution superior to the predecessor, relying on an 8-microphone array, mounted on a HATS
  - An 8-loudspeaker array is recommended for playback to achieve high robustness
  - The sweet spot not large enough [1]
  - Not common in the industry

[1] P. Klinke, R. Kostyk, J. Banas, P. Maziewski, and D. Stanczak, "Practical Evaluation of Sweet Spot in Current Noise Reproduction Systems," Engineering Brief 438, (2018 May.). doi: https://www.aes.org/e-lib/browse.cfm?elib=19551

State of the art

- Ambisonics
  - A defined method for sound capture (encoding), offering 3 dimensions, rather than just 2
  - The size of the sweet spot increases with the order of Ambisonics (degrees of freedom)
  - Sound synthesis (decoding) can be performed in various ways to match, e.g., the loudspeaker array

Recording

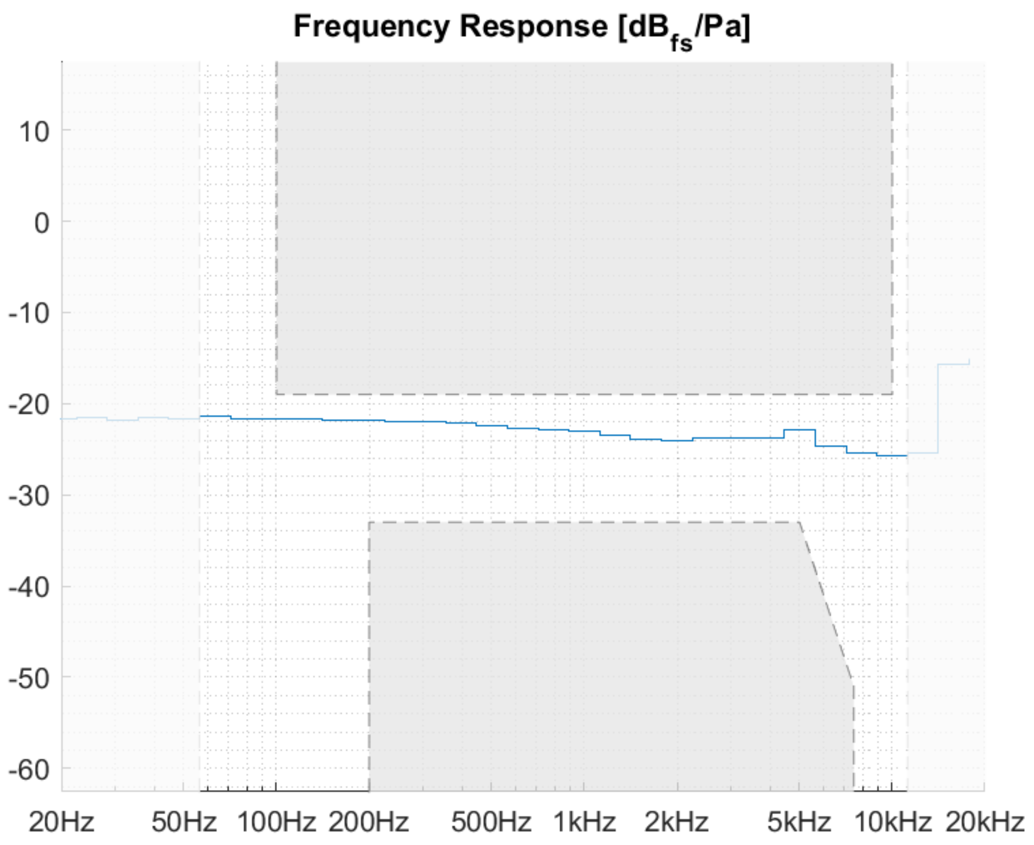
- MH Acoustics Eigenmike EM32
  - A 32-channel spherical microphone array
  - Software enabling up to 4^{th} order Ambisonics encoding
  - 3-dimensional beamforming

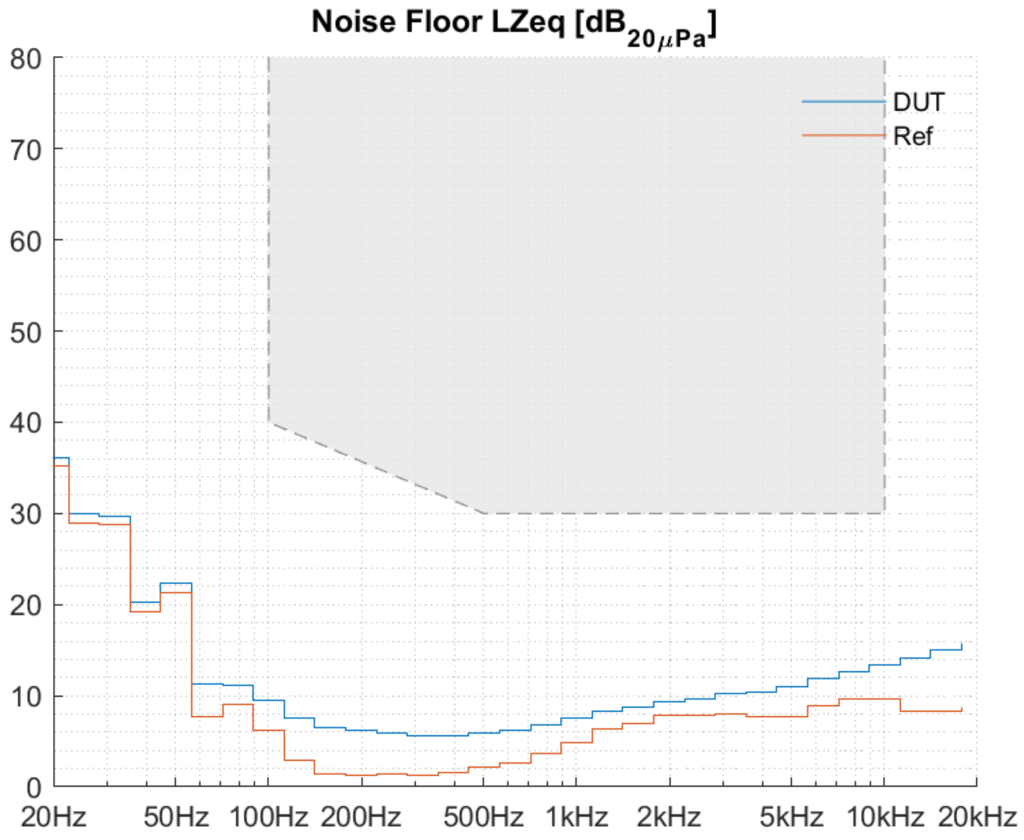
Recording

- MH Acoustics Eigenmike EM32 was tested in an anechoic chamber against Bruel & Kjaer 4189 reference class 1 microphone
- Frequency response, noise floor and harmonic distortion was measured

STQ Workshop

Recording





Recreating Complex Soundscapes for Audio Quality Evaluation by Jan Banas (Intel Corporation)

Recording

- Lifelike environments

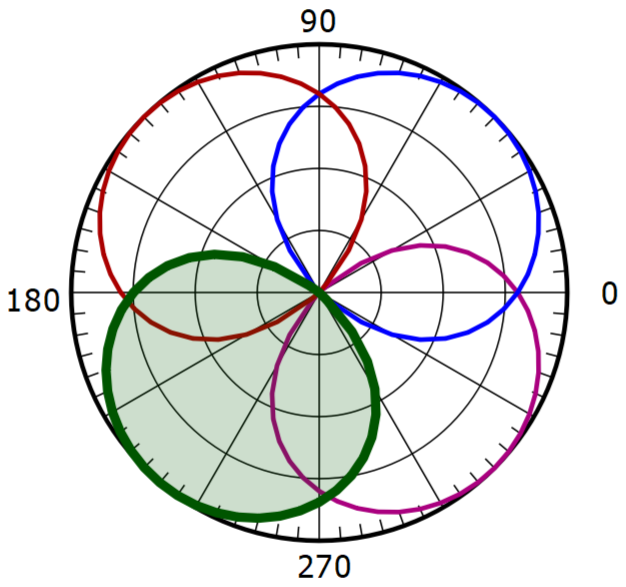


Processing

- Based on our knowledge of the industry, we decided to aim for the 4-loudspeaker array, similar to one recommended in ES 202 396-1
- Consequently, only 2 dimensions are used
- The Regular Polygon Decoder was utilized, effectively simulating 4 virtual microphones

Processing

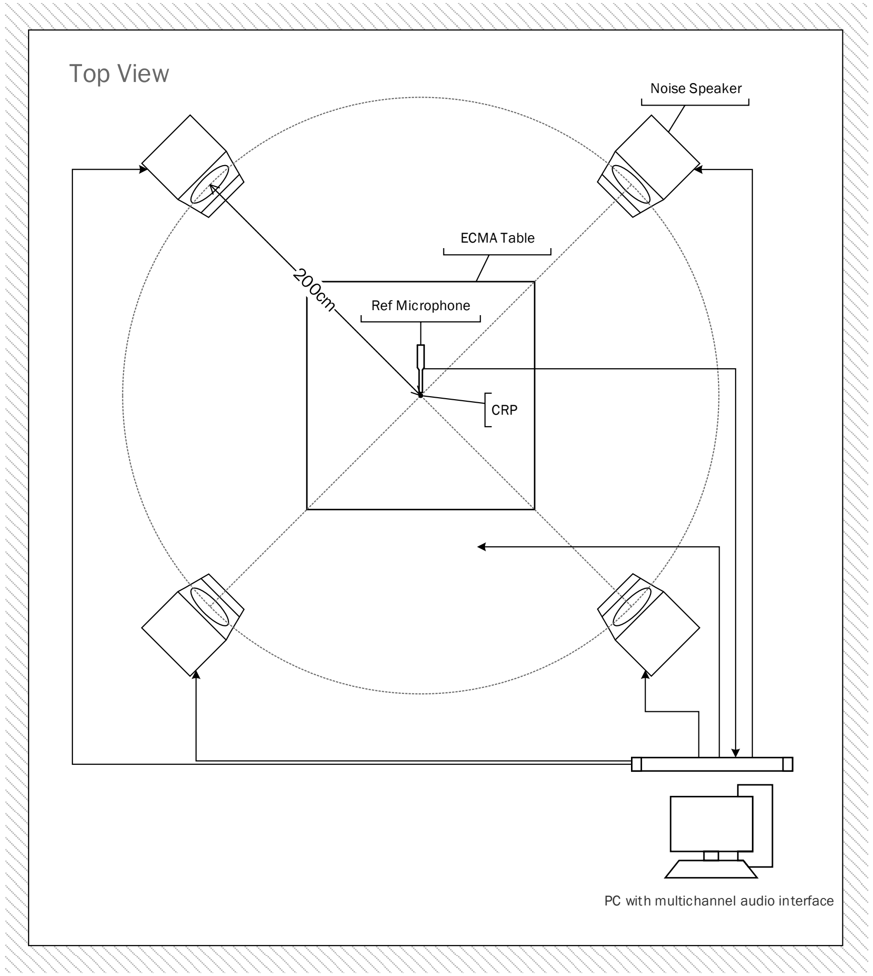
- B-Format decoding to 4 channels was achieved:
  - $LF = \left(\sqrt{2}W + X + Y\right)\sqrt{8}$
  - $LB = \left(\sqrt{2}W - X + Y\right)\sqrt{8}$
  - $RB = \left(\sqrt{2}W - X - Y\right)\sqrt{8}$
  - $RF = \left(\sqrt{2}W + X - Y\right)\sqrt{8}$



STQ Workshop

Playback

- Compliance with ES 202 396-1:
  - 4 loudspeakers equally spanned on a circle
  - ECMA Table in the middle to put a Device Under Test

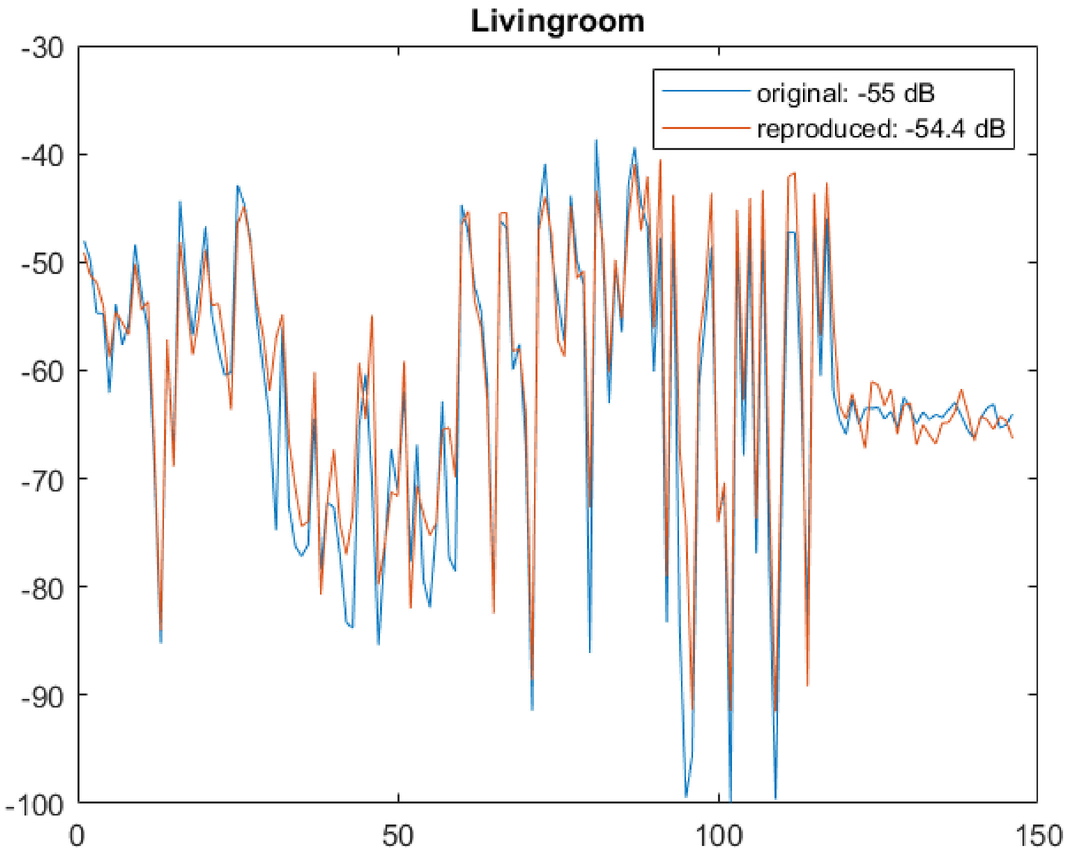


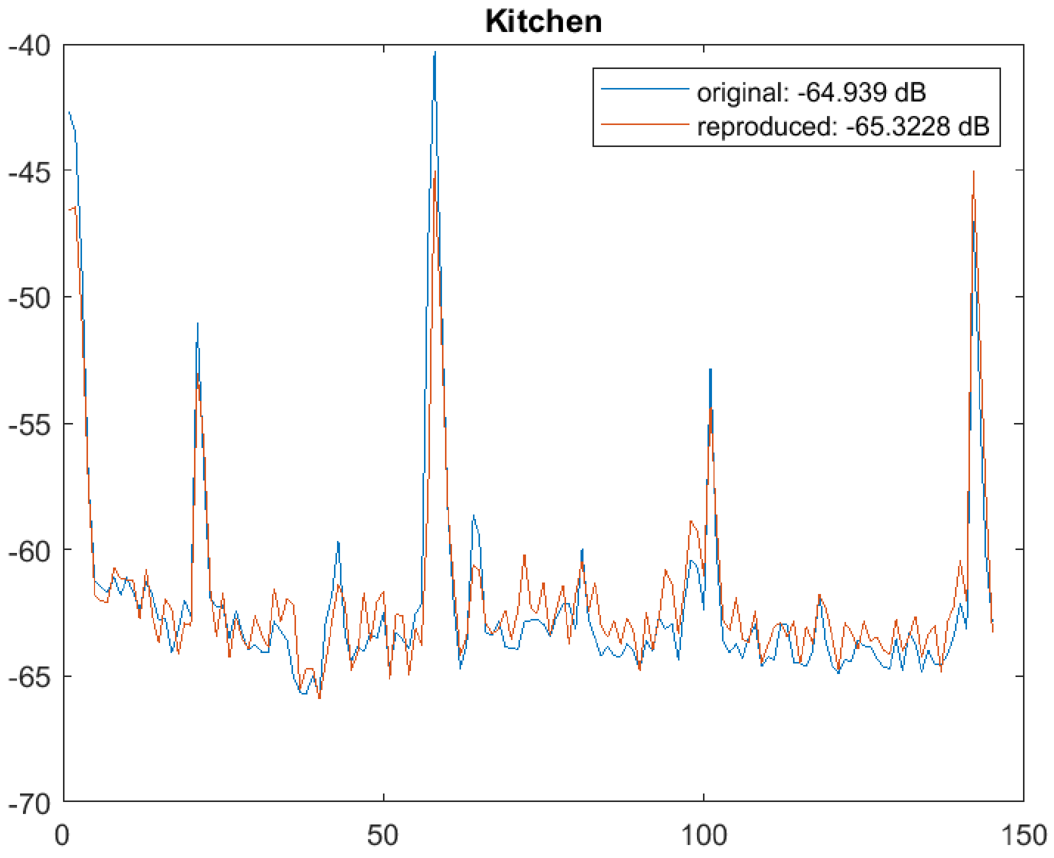
Playback

- Calibrated with pink noise using a single reference microphone
- Loudspeakers equalized to exhibit +/- 0.5dB flat frequency response in 1/3rd octave sub-bands between 100Hz and 8000Hz
- Calibration was verified by computing segmental FS level in 500ms-long segments

STQ Workshop

Playback





Conclusion

- The proposed solution allows for recording and recreating dynamic and transient-rich sound fields
- The synthesis stage is limited to the most widespread loudspeaker layout, but the source material could be decoded up to the 4^{th} order Ambisonics signals
- Compared to the state-of-the-art systems, this approach offers more simplicity and accessibility
- The sweet spot area is yet to be determined

Thank you!

- Contact us at:
  - itp.audio.lab.team@intel.com
  - jan.banas@intel.com
  - pawel.trella@intel.com
  - damian.koszewski@intel.com