MPEG-D SPATIAL AUDIO OBJECT CODING FOR DIALOGUE ENHANCEMENT (SAOC-DE)


Fraunhofer IIS
OUTLINE

- Introduction to Spatial Audio Object Coding
- Dialogue enhancement
  - Concept
  - Technical solution
- Evaluation: subjective listening test
- Demonstration
- Conclusions
Spatial Audio Object Coding (SAOC)

- SAOC is an efficient way for coding audio objects
- A semantic pre-/post-processor to other audio codecs
Dialogue Enhancement (DE) – Legacy decoder
Dialogue Enhancement (DE) – SAOC-DE decoder
Wimbledon 2011 experiment with BBC

- Public test during Tennis Grand Slam Championships 2011 in Wimbledon
- Player provided on BBC website
- User control for dialogue vs. background balance
- Well-received: >80% of people answering the questionnaire indicated the effect of this functionality to be positive
Dialogue enhancement functionality benefits

- Improving intelligibility of dialogue by background attenuation
  - Hearing-impaired audience
  - Non-native audience
  - Noisy listening environments
- Reducing dialogue level
  - Custom mixture, e.g., focusing on sports event atmosphere sounds
- Allows addressing complaints in broadcast regarding dialogue mixing levels
Adapting MPEG-D SAOC for DE-application needs

- Adding support for more than 2 downmix channels
  - In 5.1, dialogue is normally present in the 3 front channels
- Disable functionality not needed by the application
  - Reduction in engineering and computational complexity

- 2 meta-objects: foreground (FGO), background (BGO)
- Restrict rendering functionality to in-place re-balancing of FGO and BGO
  - Number of output channels = number of downmix channels
  - No object re-panning
SAOC-DE decoder overview
Signal model

- Downmix signal obtained with instantaneous linear mixing
  \[ X = DS = X_{BGO} + X_{FGO} \]

- Target output signal: re-balanced mix of partial downmixes
  \[ \hat{X} \approx m_{BGO} X_{BGO} + m_{FGO} X_{FGO} \]

- The gains determined from single user input \( m_G \)
  - The re-balancing gain to be applied to FGO (dialogue)
  - Applied as attenuation only to avoid clipping
Object reconstruction

- Parametric object reconstruction
  \[ \hat{S} = G X \approx S \]

- Un-mixing matrix
  \[ G = ED^* \left( DED^* \right)^{-1} \]

- Object covariance \( E \) and downmixing matrix \( D \) obtained from SAOC side information (object level differences, inter-object correlations, etc.)

- Enhanced Audio Object (EAO) reconstruction
  - Include waveform residual signal(s) for improving perceptual quality of output
  - Residual signals focused on FGO (i.e., dialogue signal)
    \[ X_{res} = S_{FGO} - \hat{S}_{FGO} \]
Modification Range Control (MRC)

- Rendering gain interface allows full separation of partial downmixes of FGO and BGO
  - Rights-management issues
  - Quality control concerns
- MRC values in bitstream restrict the range of the user input gain
- Content-provider retains control over the allowed modifications
  - Keep modification in a safe region
Main differences from classic MPEG-D SAOC

- Increased maximum number of downmix channels to 3 (from 2)
  - Generalized object reconstruction algorithms to support higher number of downmix channels
- Simplified rendering interface
  - In-place gain change of two meta-objects
  - Single gain rendering control input instead of a full rendering matrix
- Significant complexity reduction
  - Disabled un-used tools and modes (e.g., decorrelators)
  - Limited number of objects to 6 (from 32)
- Replaced DCU with MRC
- Removed MPEG Surround transcoding
Subjective listening test with MUSHRA

- Material similar to broadcast content
  - Stereo background (music, sound effects, audience noise)
  - Mono dialogue foreground panned to center
  - Downmix (stereo) SNR 1.5 – 7.5 dB

<table>
<thead>
<tr>
<th>Conditions</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>HR</td>
<td>Original objects mixed with target gains</td>
</tr>
<tr>
<td>HE-AAC REF</td>
<td>“HR” encoded with HE-AAC at 64 kbps</td>
</tr>
<tr>
<td>HE-AAC MO</td>
<td>Multi-Object transmission (discrete objects, 43+21=64 kbps)</td>
</tr>
<tr>
<td>HE-AAC DE</td>
<td>SAOC-DE and HE-AAC encoded downmix (64 kbps)</td>
</tr>
<tr>
<td>HE-AAC DE+</td>
<td>SAOC-DE bitrate on top of dmx bitrate (64+SAOC kbps)</td>
</tr>
<tr>
<td>LP35</td>
<td>3.5 kHz low-pass anchor</td>
</tr>
</tbody>
</table>
FGO +6 dB (parametric-only, 10 listeners, 64 kbps)
FGO +12 dB (with EAO processing, 10 listeners, 64 kbps)
Demonstration
Conclusions

- SAOC-DE extends MPEG-D SAOC for dialogue enhancement applications
- It is answering the request from broadcasting industry
  - Providing a backward compatible extension to existing services
- Good performance in subjective evaluations
- Technical solution standardized as an amendment to MPEG-D SAOC
- Included as “Advanced Clean Audio” solution in DVB
  - ETSI TS 101 154 v2.2.1