ITU-T Workshop on
"From Speech to Audio: bandwidth extension, binaural perception"
Lannion, France, 10-12 September 2008

Creation of test material that simulates the stereo capture of a teleconference site

Gunilla Berndtsson,
Senior research engineer,
Ericsson Research

Audio and video conferencing

- Key application for stereo audio codecs
- Important to have good test material that is representative of teleconferencing sessions
  - Audio scene properties
  - Stereo image capturing
  - Suitable test material

Audio scene in a conference

- Key spatial audio components:
  - Reverberated signals from the main speakers in the room
  - Background noise containing both diffused components and spatially placed components such as interfering talkers

Stereo image capturing

- Good sound quality is not enough.
- The spatial audio scene must be captured.
- A listener should be able to spatially separate the different sound sources.

Stereo capture methods

- Several stereo capture techniques available. Usual methods: A-B, Mid-Side (M-S), X-Y, near coincident (ORTF) and binaural capturing.
- Binaural stereo capture is intended for binaural rendering through headphones - not very suitable for loudspeaker rendering
- Near coincident techniques - more convincing stereo image when played back over properly placed loudspeakers, than in headphones.
- The A-B, M-S and X-Y techniques should work for both loudspeaker and headphone delivery.

Audio and video conferencing with stereo audio

- In a video conference supporting stereo audio the rendered video scene and the rendered audio scene must correlate.
- In a pure audio teleconference there is more flexibility in constructing the audio scene (microphone placement, SNR, critical length, etc)

Audio scenes for codec evaluation

- The scene should have several speakers spaced from the far left to far right.
- Each speaker should say something long enough for the listener to estimate the location of that speaker in the scene.
- Other interresting spatial aspects are simultaneous speakers at different locations, moving speakers, and typical conference background noises.

Live recording with background noise

- Advantages
  - The speakers automatically adapt to the background level in the room (The Lombard effect, voice spectrum slope).
  - Easy to obtain typical and correct background noises.
  - The radiation of sound from the sound sources is correct.
- Disadvantages
  - No possibility to create different audio scenes from the recorded material.

Audio scene simulation

- Record sentences in an anechoic room
- Use room impulse responses for the different chairs in the conference room for simulating a speaker in a particular chair.
- Interfering talkers can also be recorded in an anechoic room and be positioned using the room IR.
- Record live office noise in the same room as the room impulse responses were recorded in.
- Mix the speech with the background noise.

Audio scene simulation

- Advantages
  - Flexible (placement, SNR,...).
- Disadvantages
  - Not always as realistic as live recording.

Recent Test Example

- The Qualification Quality Assessment Test for the joint SWB extension of G.718 and G.729.1 (ITU-R BS.1285 method, headphone presentation).
  - Used scenarios:
    - 1) Large audio conference room, 12 participants sitting round a table.
    - 2) Small video conference room, 7 participants in front of a screen.

The Scene Simulation

- Each stereo sample simulated an audio scene with two talkers, a single sentence per talker.
- The talkers were placed in different positions using room impulse responses.
- For clean speech conditions the test samples were 6 seconds long, some including overlapping speech.
- For background noise conditions, the test samples were 7 seconds long with no overlapping speech.

Issues for further Exploration

- Relevant test questions for spatial quality
- Common words to describe spatial quality
- Binaural perception
- Cognitive functions
- User expectations
- The quality of teleconferencing today
- How will teleconferencing be used?
- How disturbing are different background noises?
- How to achieve good SNR in conferencing?

- Thank you!