Prototype for IETF Interactive Audio Codec

draft-valin-codec-prototype-01.txt

Jean-Marc Valin, Octasic Inc.
Koen Vos, Skype Technologies S.A.

IETF 78 – codec WG
Introduction

• Lots of technical work since last meeting
• This is a prototype, not a complete codec
  • No decision is final
• Aligned on draft-ietf-codec-requirements-00
• Hybrid codec based on a combination of the SILK and CELT codecs
  • SILK: speech frequencies below 8 kHz
  • CELT: music and speech frequencies above 8 kHz
  • Other codecs not excluded
Work Since Last Meeting

• SILK pre-integration work
  • Conversion to use the CELT range coder
  • Support for 10 ms frames
  • Resampling for 32 kHz and 48 kHz sampling rates

• CELT pre-integration work
  • Optimisation of 20 ms frames, dynamic switching
  • Adjustable frequency range (low and high)
  • Tuning, quality improvements

• Actual integration work
Where is the Secret Source Code?

- All development is publicly accessible with git
- Main CELT repository
  - git://git.xiph.org/celt.git
- Temporary (unofficial) SILK repository
  - git://git.xiph.org/users/jm/silk.git
- Jean-Marc's version of the hybrid codec
  - git://git.xiph.org/users/jm/ietfcodec.git
Characteristics

- Sampling rates: 8, 12, 16, 32, 48 kHz
- Audio bandwidth: 4, 6, 8, 16, 20 kHz
  - Narrowband, medium band, wideband, superwideband, fullband
- Bitrates: ~6 – 128 kb/s per channel
- Frame sizes: 2.5, 5, 10, 20, 40, 60 ms
- Look-ahead: 2.5 – 5.2 ms
- Audio contents: speech, music
Operating Modes

- SILK-only
  - wideband: 10, 20, 40, 60 ms
  - medium band: 10, 20, 40, 60 ms
  - narrowband: 10, 20, 40, 60 ms

- SILK+CELT (hybrid)
  - fullband: 10, 20 ms
  - superwideband: 10, 20 ms

- CELT-only
  - fullband: 2.5, 5, 10, 20 ms
  - superwideband: 2.5, 5, 10, 20 ms
  - wideband: 2.5, 5, 10, 20 ms
  - narrowband: 2.5, 5, 10, 20 ms
Listening Test

Mushra results 32 kbps speech (8 listeners)
Conclusion

• Prototype codec for speech and music
• Designed to match draft-ietf-codec-requirements-00
• Future integration work
  • Automatic mode selection/switching
  • Better hybrid mode bit allocation
  • Closer PLC integration
  • Lots of tuning
  • More code sharing