

**Proposed (DRAFT)
RTP Profile
for the
Speex Codec**

**Jean-Marc Valin
Simon Morlat
Roger Hardiman
Greg Herlein**

**IETF
Audio/Video Transport Working Group
March 20, 2003
San Francisco, CA**

About the Authors

- **Jean-Marc Valin**
 - **Codec author**
- **Simon Morlat**
 - **Linphone author (SIP Client)**
- **Roger Hardiman**
 - **OpenH323 Contributor**
- **Greg Herlein**
 - **Cat Herder, Profile Editor**

About Speex

Speex is based on the CELP encoding technique with support for narrowband (nominal 8kHz), wideband (nominal 16kHz) or ultrawideband (nominal 32kHz) sampling.

- **Open-source implementation**
- **Patent and royalty-free**
- **Integration of wideband and narrowband in the same bit-stream**
- **Dynamic bit-rate switching, variable bit-rate (VBR). Average bit-rate (ABR) supported by adjusting quality in VBR**

More About Speex

- **Voice Activity Detection (VAD) and Discontinuous Transmission (DTX)**
- **Support for Packet Loss Concealment**
- **Adjustable Complexity**
- **Multiple bit-rates supported**
 - **2.15 kbps**
 - **3.95 kbps**
 - **5.95 kbps**
 - **8 kbps**
 - **11 kbps**
 - **15 kbps**
 - **18.2 kbps**
 - **24.6 kbps**
 - **and up to 44 kbps in wideband modes**

Deployment

Speex is currently in use in

- **Linphone**
- **OpenH323 - GnomeMeeting**
- **Asterisk**

Strong interest has been expressed from game developers

Work is in progress for a fixed point implementation as well as several DSP implementations

Codec Basics (Narrowband)

(values for an 8 kbps nominal bit-rate)

- **20 ms Frame Length**
- **10 ms lookahead required**
- **~10 mflops nominal encoding complexity**
- **~1 mflops decoding complexity**

Codec Basics (Wideband)

(values for a 20.6 kbps nominal bit-rate)

- **20 ms Frame Length**
- **14 ms lookahead required**
- **~19 mflops nominal encoding complexity**
- **~2 mflops decoding complexity**

Codec Basics (Ultrawideband)

(values for an 22.4 kbps nominal bit-rate)

- **20 ms Frame Length**
- **16 ms lookahead required**
- **~25 mflops nominal encoding complexity**
- **~4 mflops nominal decoding complexity**

Bit-stream Characteristics

- **Supports mixed-mode operation (wideband and narrowband in the same bitstream)**
- **Supports dynamic bit-rate switching (either for VBR or via request) at any 20ms frame boundary**
- **Mode is included in each frame, so packet loss during a mode switch does not affect performance**
- **Decoder is downward compatible**
(less capable decoder can recover what bands it supports with no degradation in that band – it discards the band data that is incapable of handling)

Speex's In-band Signaling

(transparent using libspeex - currently treated as suggestions only - still subject to change if the IETF has good ideas)

- **Decoder Perceptual Enhancement Request**
- **Encoder Mode Switch Request**
- **Encoder Mode Switch Request Low Band**
- **Encoder Mode Switch Request High Band**
- **Encoder Quality Switch Request for VBR**
- **Request Acknowledge**
- **Encoder VBR/VAD/DTX Enable Request**
- **Transmit 8-bit Character**
- **Intensity Stereo Information**
- **Announce Maximum Acceptable Bit-Rate**
- **Packet N Receive Acknowledge**
- **Conservative Encoding Request**
- **User Defined**

RTP Profile Requirements

To use Speex in RTP, a profile must provide the following:

- **Mechanism for Dynamic Payload Type Mapping**
- **Mechanism to indicate when more than the default one codec frame per packet is sent**
- **Mechanism to indicate the sample rate of the encoded bitstream for accurate playback**
- **Mechanism to indicate the decoders desired bandwidth, CBR or VBR, and Comfort Noise Generation.**

Draft Profile

- **Mechanism for Dynamic Payload Type Mapping**

SDP example:

**m=audio 8088 RTP/AVP 97
a=rtpmap:97 speex/8000**

H.323 Example:

**Non-standard codec block definition in the H.245 codec
capability negotiations:**

t35CountryCode = Hex: B5

t35Extension = Hex: 00

**manufacturerCode = Hex: 0026 Length of the Binary
Sequence (8 bit) Binary Sequence consisting of an
ASCII string, no NULL terminator**

Draft Profile

- **Mechanism to indicate when more than the default one codec frame per packet is sent**

RECOMMENDED practice is to use only one frame per packet.

ptime parameter set to increments of 20ms to indicate the number of Speex frames per packet.

Draft Profile

SDP Example:

```
m=audio 8008 RTP/AVP 97  
a=rtpmap:97 speex/8000  
a=ptime:40
```

H.323 Example:

Append the string “ptime:40” to the binary sequence

Draft Profile

- Mechanism to indicate the sample rate of the encoded bitstream for accurate playback

Typically defined in the *rtpmap* parameter, but may also be specified using the *sr* parameter

SDP Example:

```
m=audio 8088 RTP/AVP 97  
a=rtpmap:97 speex/8000
```

H.323 Example:

append the string “sr:8000” to the binary sequence

Draft Profile

- **Mechanism to indicate the decoders desired bandwidth, CBR or VBR, and Comfort Noise Generation.**

Encoding Bandwidth

ebw: narrow|wide|ultra

Variable Bit Rate

vbr: on|off

Comfort Noise Generation

cng: on|off

Mode

mode: <numeral>

Perceptual Enhancement

penh: on|off

Known Issues

- **MIME Type Registration Required**
- ***ptime* statement should read that default is 20 (one frame), not one.**
- **Does the initial profile need to cover stereo?**
- **Current libspeex Library is stable at 1.0rc3 and is ready for full 1.0 release, pending IETF comments to the contrary**
- **Others?**

Conclusion

Because Speex is patent and license Free, has good performance, and is available as an Open Source library under a liberal BSD-style license, we believe that Speex will radically lower the barrier for entry for speech compression in next-generation network applications.

The next step is to standardize how Speex frames are encoded onto the wire.

This Draft Profile is a beginning of that step, and we thank you for considering it.