Speex: Speech Compression for Everyone

Jean-Marc Valin
Ph.D. Candidate
University of Sherbrooke

Overview

- What is Speex?
- Introduction to CELP
- Features
- Narrowband, wideband
- Applications
- Standards
- What's next?

What is Speex?

- Audio codec optimized for speech and VoIP
  - Started in February 2002
  - Bit-stream frozen as of March 2003
- Open-Source/Free Software (BSD)
- Patent-free
- Part of the Ogg project (Xiph.org foundation)
- Part of the GNU project
- Not related to the Sherbrooke speech coding lab

Project goals

- Provide an alternative to expensive, proprietary speech codecs (e.g. G.723.1, G.729)
- Help bring good quality VoIP to Linux and other Free OS
- Lower barrier of entry for VoIP in general
- Allow a wide range of speech applications
  - Both on-disk and over network
  - Wide range of bit-rates, complexity

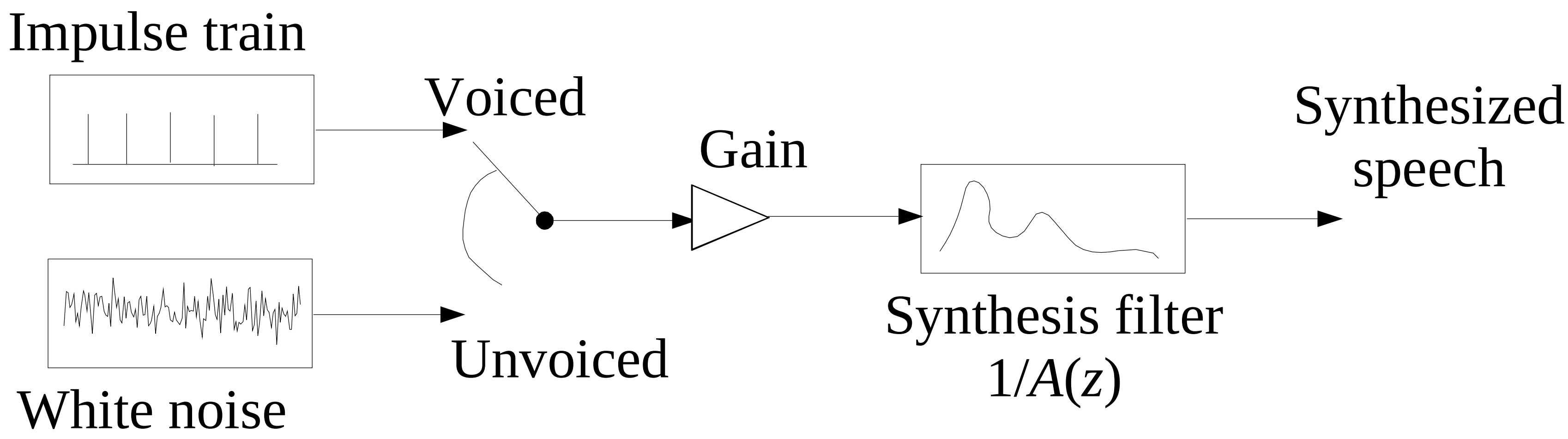
Design requirements

- Integrate multiple sampling rates and bit-rates
- Portable code, reasonable complexity
- Be robust to packet loss (VoIP operation)
- Best quality possible given current patent status
  - It's not the best speech codec out there, but it's close enough

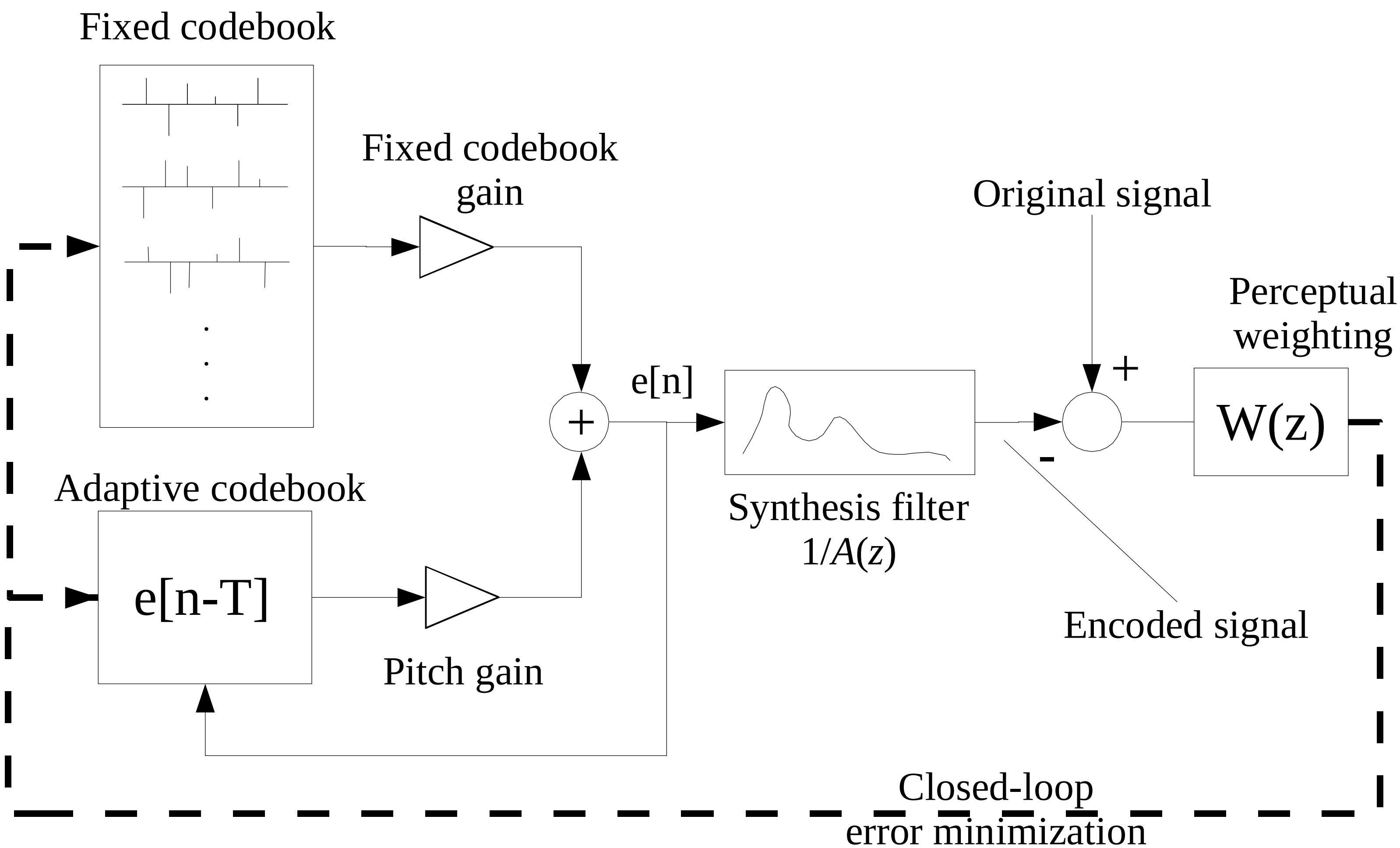
What does patent-free mean exactly?

- Strategies against patents
  - Make sure to avoid known patents (e.g. ACELP)
  - Search USPTO site
  - Use technology from the 80's
- No absolute warranty regarding patents
  - Patents are vague, some are not yet disclosed
  - Different in many countries
  - Even a danger for proprietary codecs/formats

Speech production model



CELP – Code-Excited Linear Prediction



Main features

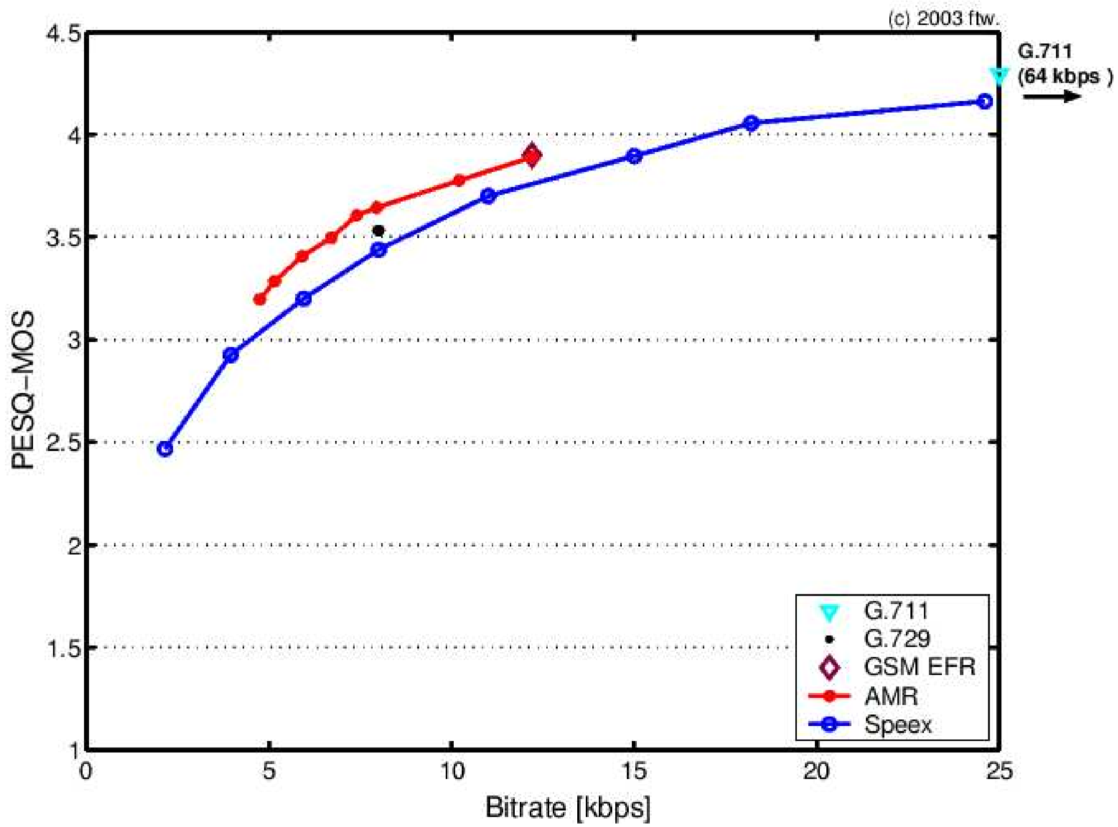
- CELP-based
- 20 ms frame size (+10 ms look-ahead)
- Narrowband and wideband in the same bit-rate
- Bit-rates ranging from 2 kbps to 44 kbps
  - Bit-rate can change dynamically
- VBR operation, VAD, DTX
- Stereo support
- Variable complexity

Narrowband Speex

- Bit-rates: 2.2 kbps to 24.6 kbps
- Recommended for VoIP: 8 kbps, 11kbps, 15 kbps
- Samples
  - Original
  - 15 kbps
  - 8 kbps
  - 4 kbps

Narrowband quality

- Evaluation using PESQ (by Florian Hammer)
  - This isn't a real MOS test



Wideband Speex

- Wideband is the future!
  - The only way to beat PSTN quality
- Speex wideband is compatible with narrowband
- Recommended for VoIP: 12.8 kbps to 27.8 kbps
- Samples
  - Original
  - 27.8 kbps
  - 20.6 kbps
  - 12.8 kbps

Writing Speex applications

- Send the right input
  - Remove DC, apply IRS filtering for narrowband
  - Use correct gain (no clipping)
- Set codec parameters
  - Sampling-rate, modes
  - VBR, ABR, VAD, DTX
  - Quality, bit-rate, complexity
  - Perceptual enhancement (decoder)
- Handle lost packets

Speex vs. Proprietary Codecs

- Pros
  - Open-source, patent-free, $$$
  - Replaces a whole family of codecs
  - Good quality
- Cons
  - Some proprietary codecs achieve better quality/bit-rate
  - Not adapted to non-IP wireless (e.g. cell phones)

Standards

- RTP draft submitted to IETF (http://www.ietf.org/internet-drafts/draft-herlein-speex-rtp-profile-00.txt)
- MIME type application to come
- Aiming for *de facto* standard
- ITU-T standardization (G.7xx)
  - Did you just find $100,000 in your back pocket?

Software support

- Linphone (first VoIP app to use Speex)
- OpenH323 (ohphone, GnomeMeeting)
- Various Plugins
  - WinAmp
  - XMMS
  - DirectShow
  - ACM
- Unreal Tournament (coming soon!)

The future

- Fixed-point implementation
  - Speex on PDA's
  - DSP applications
- Make libspeex a complete VoIP solution (echo cancelling, denoising, AGC, ...)?
- MIME type application
- Integrate both voice and music in the same codec