

MPEG-4 Speech Coding

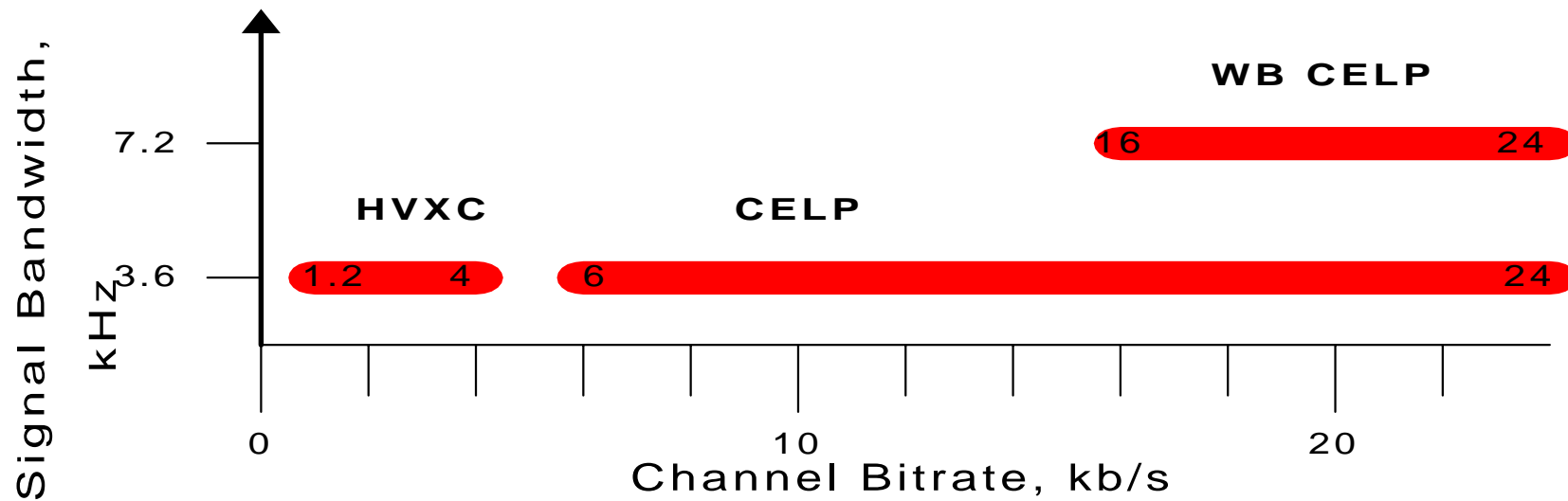
Schuyler Quackenbush
AT&T Labs - Research
Florham Park, NJ 07932 USA
srq@research.att.com

Speech Coding Overview

- Excellent compression by using source model
 - Linear Predictive Coding (LPC)
 - Pitch or noise excitation
- Better compression than “general audio” coders
 - only for “clean speech” from single talker

Speech Coders

- Harmonic Vector Excitation Coder (HVXC)
- Code Excitation Linear Prediction (CELP)
- Wideband CELP



Communication Characteristics

- Low bit rate
 - HVXC 1.2 kb/s to 1.7 kb/s var. rate
2.0 kb/s to 4.0 kb/s const. rate
 - CELP 4.0 kb/s to 24 kb/s const. Rate
- Low one-way delay
 - HVXC 33.5 ms to 56 ms
 - CELP 15 ms to 45 ms
- Not compromised for modem signals

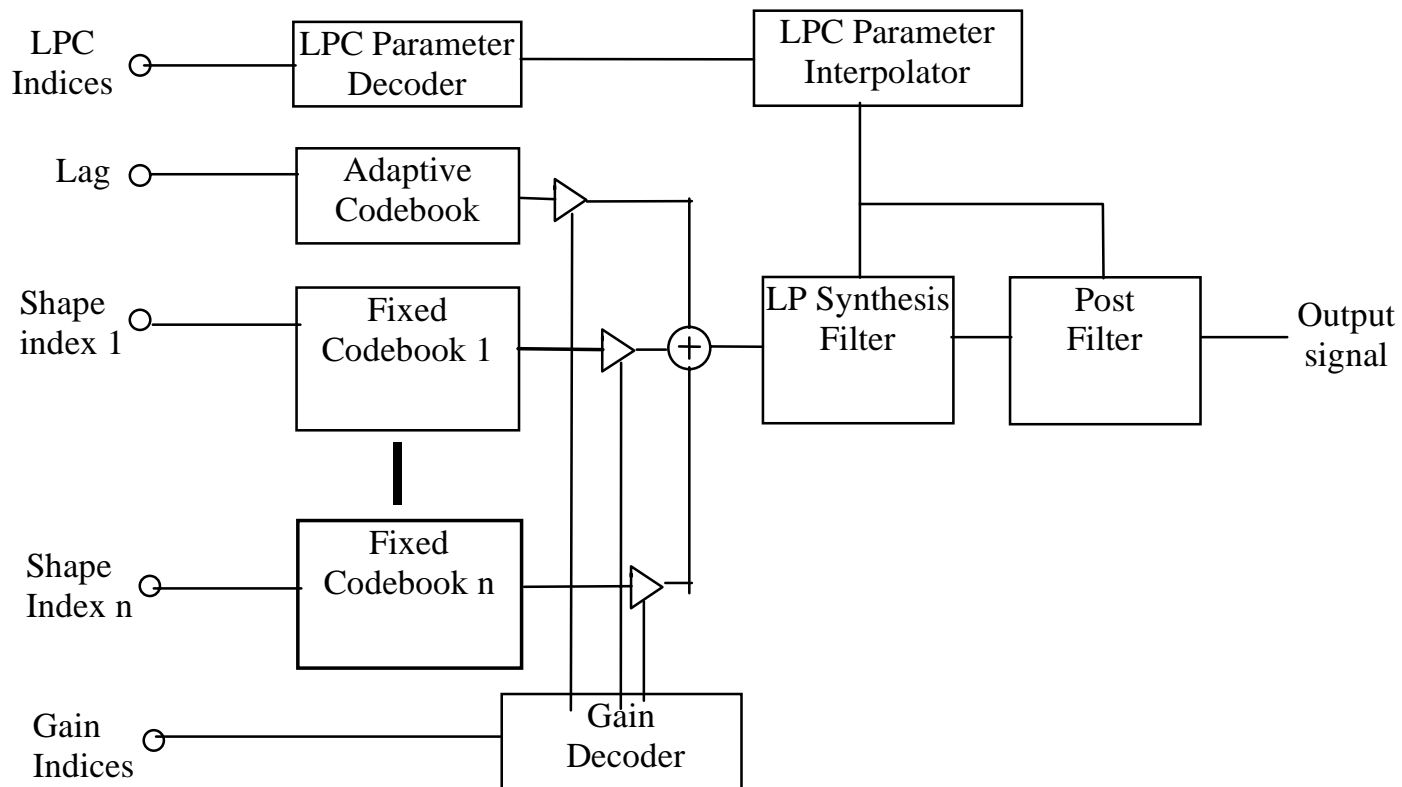
Bit Rate Scalability

- Parameters coded using multi-stage VQ
 - base plus enhancement layer
- Enhancement layers can be stripped in
 - server
 - channel
 - decoder



Parameter Update Scalability

- Frames are 10 ms to 40 ms
 - 2 to 10 sub-frames per frame
- Linear Prediction Model
 - updated every frame
 - interpolated every sub-frame
- Excitation
 - gain updated every subframe

CELP Decoder Block Diagram



Demonstration

- 1.7 kb/s variable rate HVXC 
- 12 kb/s CELP 
- 24 kb/s Wideband CELP 