

MPEG Unified Speech and Audio Coding

Schuyler Quackenbush

Audio Research Labs

srq@audioresearchlabs.com

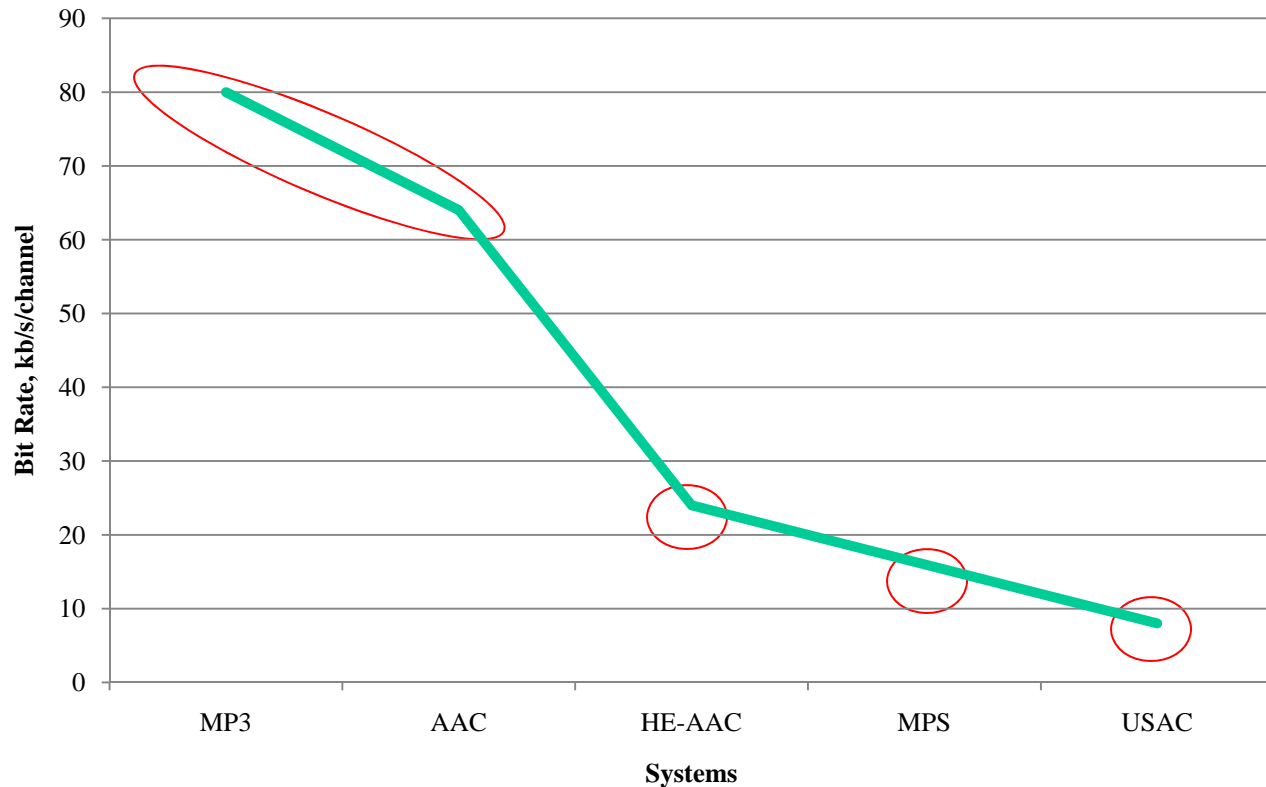
ARL

audio research labs

“Dimensions” of MPEG Audio Compression

- Perception Models
 - SNR
 - Perceptually shaped quantization noise
 - MPEG-1 Layer III (MP3); MPEG-2 Advanced Audio Coding (AAC)
 - Frequency
 - Bandwidth replication
 - MPEG-4 High Efficiency AAC (HE-AAC)
 - Space
 - Perceptual soundstage coding
 - MPEG Surround
- Production Models
 - Speech production
 - Unified Speech and Audio Coding (USAC)

Compression Performance

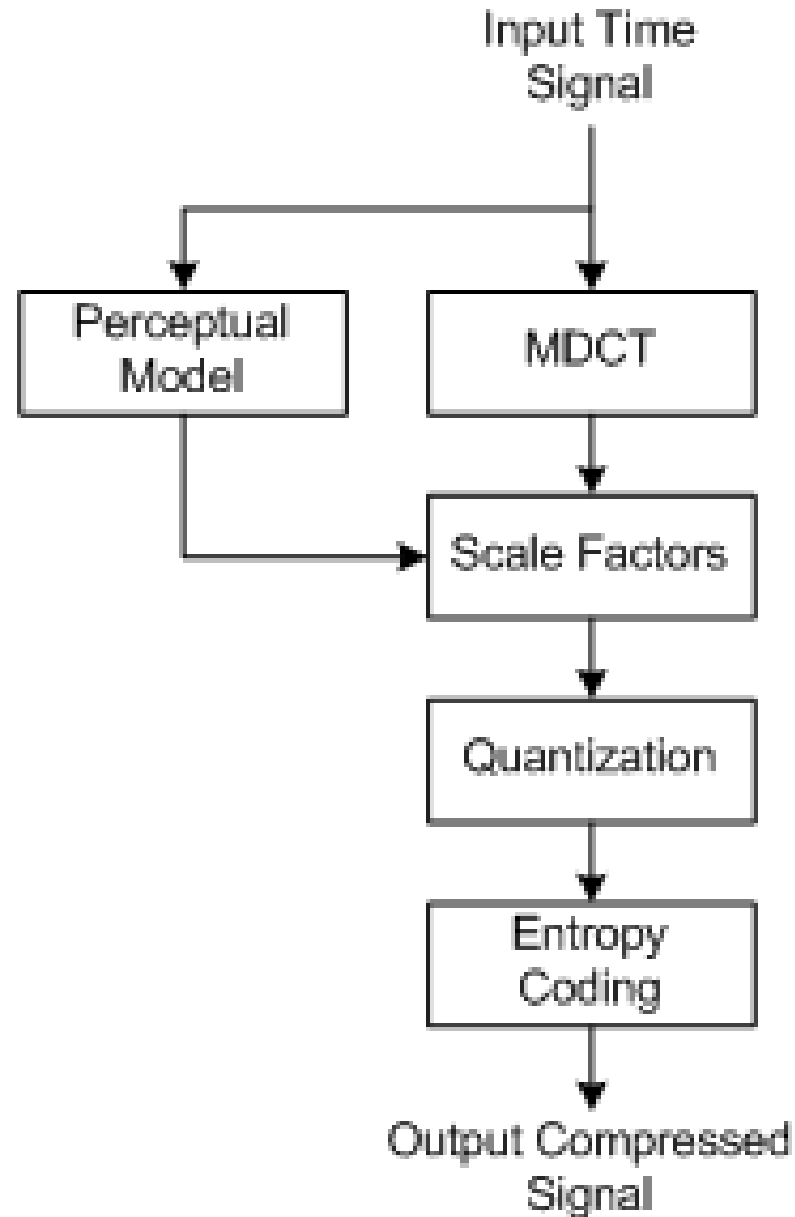


USAC Philosophy

- Adaptively select encoding mode
 - at each coding frame boundary (e.g. 20 ms)
- Coding modes are
 - Frequency Domain
 - AAC
 - TCX
 - Time Domain
 - ACELP
- Need architecture that incorporates all of these **ARL** modes!

ARL

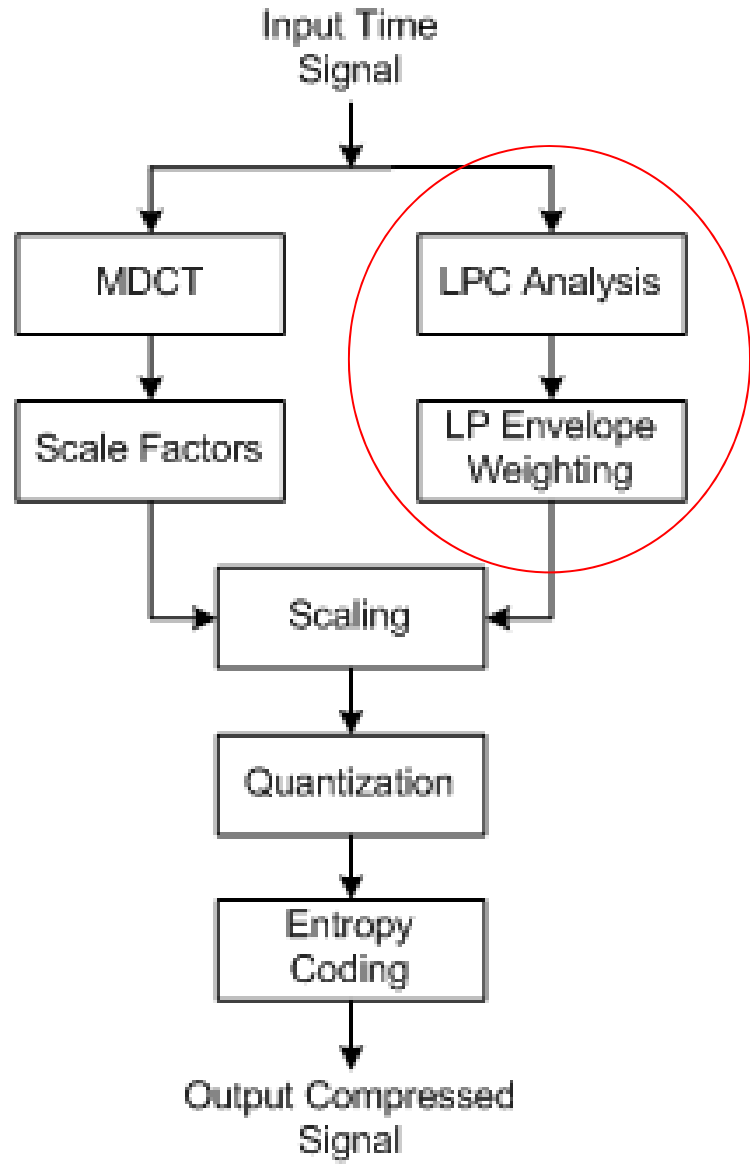
AAC Foundation



Speech Models - I

- MDCT coefficient scaling
 - Sets quantizer stepsize per psychoacoustic model
- AAC
 - Uses high-frequency resolution scalefactors
- USAC
 - Adds Transform Coded Excitation (TCX) variant to achieve medium-frequency resolution scaling via short-term LP-modeled spectrum
 - Much lower data rate!

AAC + TCX



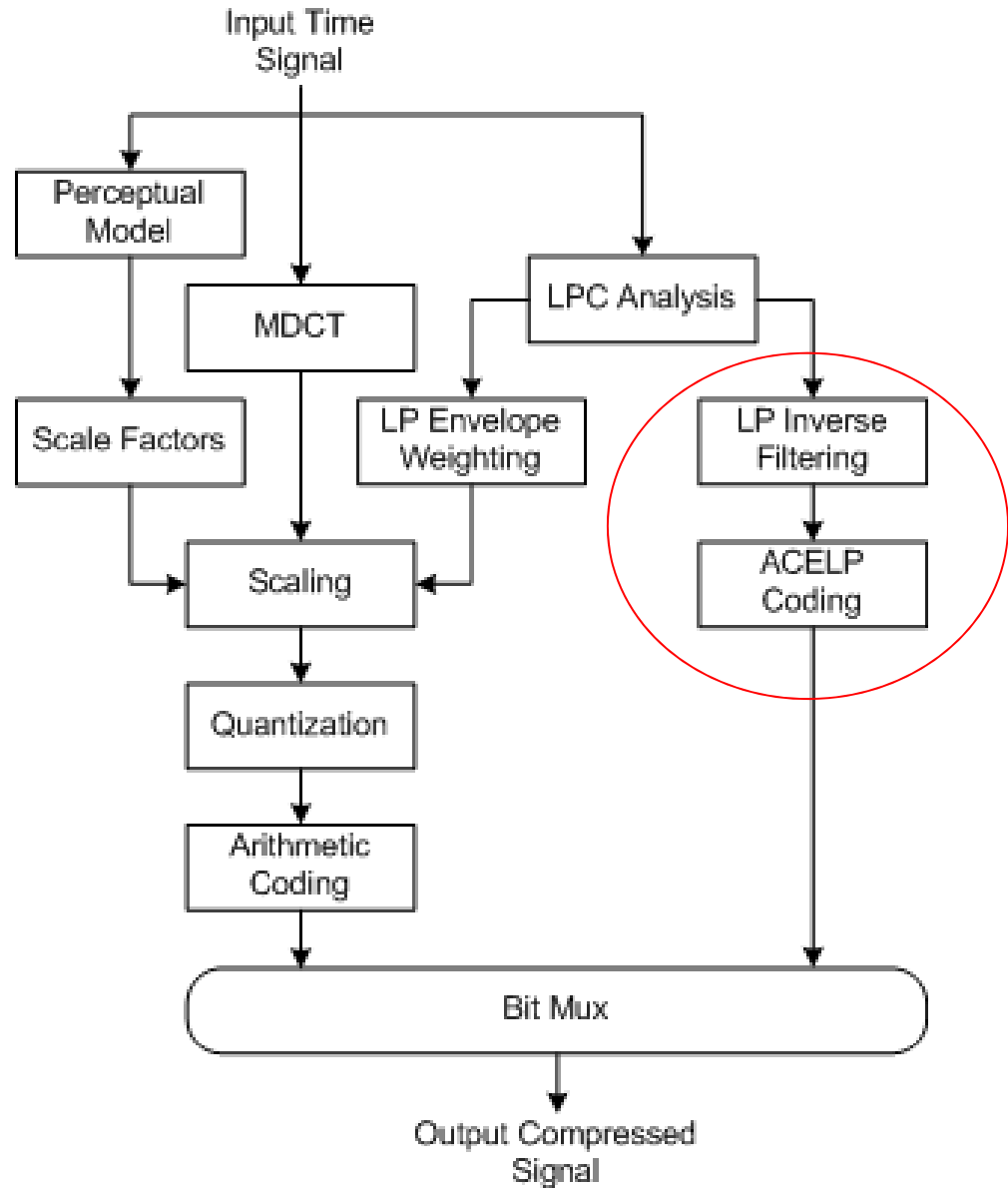
Speech Models - II

- Add time-domain Linear Prediction coding mode
 - Algebraic Coded Excitation Linear Prediction (ACELP)

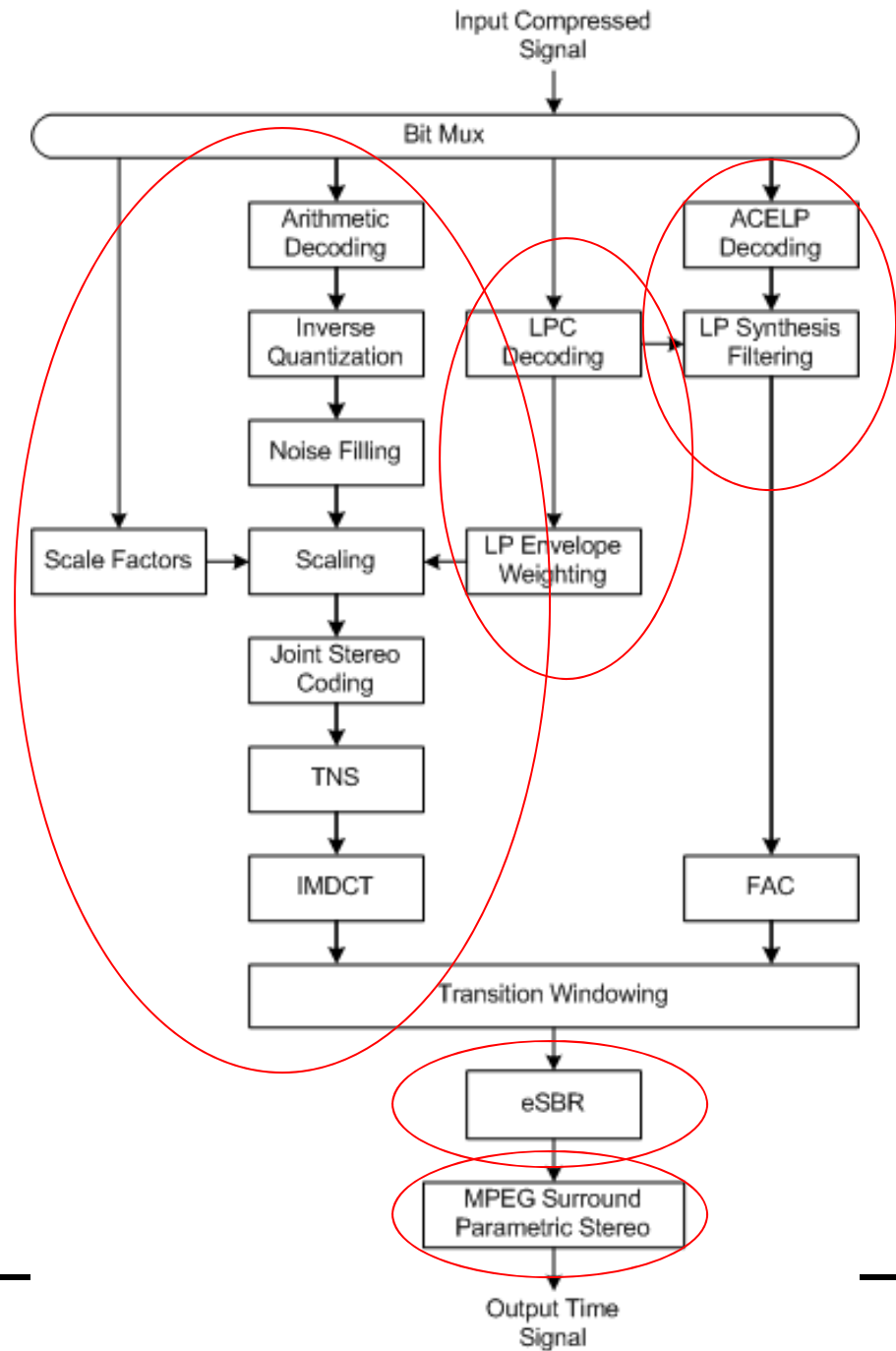
AAC+

TCX+

ACELP



USAC Decoder



Performance

ARL

audio research labs