Codec for Enhanced Voice Services (EVS)— The New 3GPP Codec for Communication

Workshop at the 140th AES Convention 2016

Stefan Bruhn, Ericsson AB
Václav Eksler, VoiceAge Corporation
Guillaume Fuchs, Fraunhofer IIS
Jon Gibbs, Huawei Technologies Co. Ltd

Introduction

- EVS Codec
  - Speech and audio codec for the next generation of (mobile) telephony and communication
  - Representation of audio content up to 20 kHz audio bandwidth
  - Designed for high quality and efficient coding of speech, music and mixed content
  - Includes high coding efficiency and enhanced packet-loss concealment for challenging channel conditions
  - → New level for user-experience for all channel conditions
- Standardization finalized in 3GPP end of 2014
- In Deployment since 2015

Workshop Overview

- Part 1: Introduction
  - 3GPP Standardization
  - General Overview over codec, features
  - Deployment
- Part 2: Performance, Application Scenarios, Demos
  - Test results and user experience
  - Demos underlying the results
- Part 3: Coding of Speech in EVS
  - Overview over speech-coding part in EVS
  - Advancements over previous standards
- Part 4: Coding of Mixed/Music Content in EVS
  - Overview over transform-coding part in EVS
  - Advancements over previous standards
- Summary, Conclusions
- Questions

Part 1: Introduction

presented by Stefan Bruhn, Ericsson AB

Part 1: Outline

- Background of the 3GPP work item
- Context of EVS within the mobile network generations
- 3GPP standardization process
- General overview over codec, block diagram, main features, operating points
- Deployment

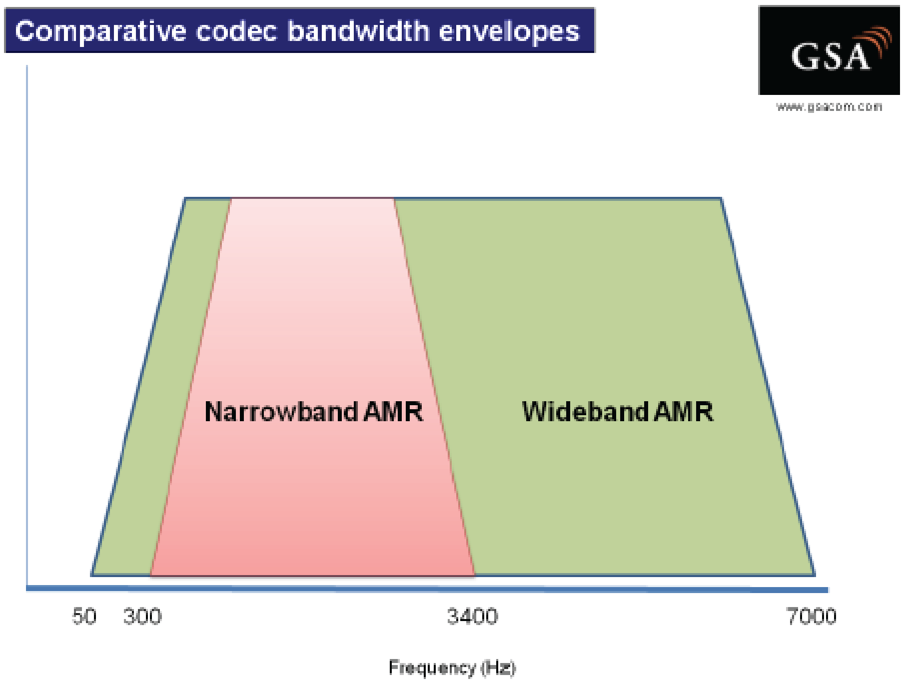
Evolution of voice Service

Traditional (narrowband) voice service

- World wide deployment in all mobile and fixed networks

HD VOICE

- 164 mobile operators commercially launched HD voice in 88 countries)*
  - 130 operators on 3G/HSPA networks
  - 17 operators on 2G/GSM networks
  - 63 operators in 35 countries on LTE networks (VoLTE HD service)
- 30% more mobile operators offering HD voice than a year ago)**
- More than 300 HD voice phones launched)**



)* GSA May 2016 http://gsacom.com/download.php?id=2987
)** GSA Sept 2014 http://www.gsacom.com/news/gsa_415.php

ERICSSON Fraunhofer IIS HUAWEI VoiceAge

HD voice - successful operator business

Orange first in the world to launch mobile HD voice in 2009
Orange HD voice launched in 17 networks

Orange studies show)* :

96% of customers are satisfied with HD Voice calls

86% of testers say compatibility with HD voice would be a selection criterion when purchasing a mobile in future

76% of testers would be prepared to change mobiles to obtain HD Voice

Further studies show that HD voice

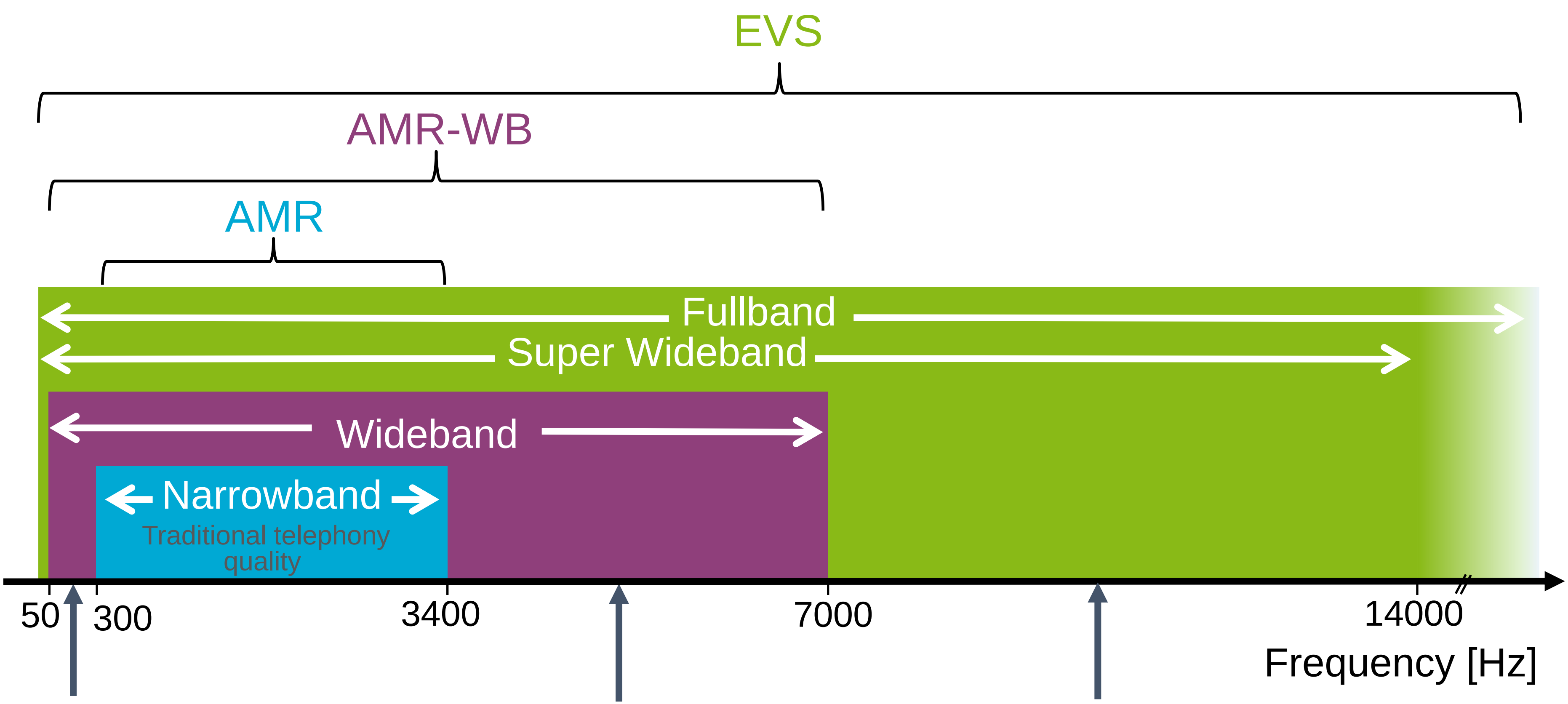
- Leads to improved user satisfaction that can turn into revenue either
  - Directly by charging (monthly fee or per minute charge), or
  - Indirectly due to reduced churn

)* GSA Nov 2015 http://www.gsacom.com/downloads/pdf/GSA_mobile_hd_voice_031115.php4

EVS – Enhanced Voice Services

- The next generation telecom voice codec for voice and music
  - For VoLTE (4G)
  - Standardization for 3G ongoing
- Standardized in 3GPP
  - Maintain 3GPP voice services cutting edge
- Next level of HD voice
  - Smooth migration from HD voice to EVS
  - Interoperability with AMR-WB

Audio bandwidth for mobile voice services



Lower Area	Mid Area	Upper Area
more natural sound, presence and comfort	More natural voice, better understanding and voice recognition	Natural sound and high quality music

3GPP Standardization

- 3GPP = 3rd Generation Partnership Project
  - 3GPP is the creator of the globally available 3G, 4G and 5G mobile communication standards
  - 3GPP unites seven telecommunications standard development organizations, the organizational partners
    - ARIB, ATIS, CCSA, ETSI, TSDI, TTA, TTC
  - 3GPP has four Technical Specifications Groups (TSG):
    - Radio Access Networks (RAN)
    - Service & Systems Aspects (SA)
    - Core Network & Terminals (CT)
    - GSM EDGE Radio Access Networks (GERAN)
  - SA WG 4 “Codec” (SA4)
    - deals with speech, audio, video, and multimedia codecs
- 3GPP TSG SA WG 4 is the creator of the EVS codec standard

Standardization Phases

- Study item phase (2007-2010)
    - Use cases and requirements for enhanced voice codecs (TR 22.813)
- Work item phase (2010-2014)
    - **Definition of Terms of Reference**
        - Design constraints
        - Performance requirements
    - **Qualification**
        - Reducing number of candidates from 13 to 5, followed by
        - **Decision to submit a single codec candidate for selection**
    - **Selection** with (upfront) agreed criteria
        - Deliverables
        - Selection rules
            - Assessing fulfilment of
                - Subjective and objective performance requirements
                - Design constraints
    - **Verification**
        - Cross-checking important parameters
    - **Characterization**
        - Evaluation of particular codec properties untested in selection
- Specification Maintenance
    - Formal 3GPP Change Request (CR) procedure

Performance Requirements

Requirements defined in relation to state-of-the-art reference codecs

- **Subjective requirements**
  - Input signal categories
    - clean speech
    - noisy speech
      - car, street, office noise
    - music and mixed content
  - VAD/DTX on/off
  - Clean and noisy channel
    - 0%, 3%, 6% FER
    - delay/loss profiles (JBM performance)
  - Input levels variations
  - AMR-WB IO in 3 interworking scenarios with legacy AMR-WB
    - AMR-WB IO encoding-AMR-WB decoding
    - AMR-WB encoding-AMR-WB IO decoding
    - AMR-WB IO encoding/decoding
- **Objective requirements**
  - Active frame rate (VAD activity)
  - Power level and inactive region attenuation
  - Maximum average bitrate (relevant for VBR)
  - JBM compliance to requirements of 3GPP TS 26.114

- **Reference codecs** standardized by 3GPP and ITU-T
  - AMR
  - AMR-WB
  - AMR-WB+
  - G.711
  - G.711.1
  - G.718
  - G.718B
  - G.719
  - G.722
  - G.722.1
  - G.722.1C

EVS Standard

- EVS core specifications
  - TS 26.441 General Overview
  - TS 26.442 ANSI C code (fixed-point)
  - TS 26.443 ANSI C code (floating-point)
  - TS 26.444 Test Sequences
  - TS 26.445 Detailed Algorithmic Description
    - Including annex with EVS RTP payload format
  - TS 26.446 AMR-WB Backward Compatible Functions
  - TS 26.447 Error Concealment of Lost Packets
  - TS 26.448 Jitter Buffer Management
  - TS 26.449 Comfort Noise Generation (CNG) Aspects
  - TS 26.450 Discontinuous Transmission (DTX)
  - TS 26.451 Voice Activity Detection (VAD)
  - TR 26.952 EVS Codec Performance Characterization

A GLOBAL INITIATIVE

- 3GPP system specifications

  PS networks (4G – LTE)
  - TS 26.114 IP Multimedia Subsystem (IMS); Multimedia telephony; Media handling and interaction
    - Mandating EVS for SWB and FB speech service, recommending EVS for NB and WB

  CS networks (3G – UMTS)
  - TS 26.453 Speech codec frame structure
  - TS 26.454 Interface to Iu, Uu, Nb and Mb

- Useful link: www.3gpp.org/sa4

- GSMA
  - PRD IR.92 VoLTE
    - Mirroring EVS status in 26.114: Mandating EVS for SWB and FB speech service, recommending EVS for NB and WB

ERICSSON Fraunhofer IIS HUAWEI VoiceAge www.voiceage.com

Range of Operating Points

Band- width	Bitrates [kbps]											
FB 20 kHz						16.4	24.4	32.0	48.0	64.0	96.0	128.0
SWB ≥ 14 kHz				9.6	13.2	16.4	24.4	32.0	48.0	64.0	96.0	128.0
WB 8 kHz	5.9 VBR	7.2	8.0	9.6	13.2	16.4	24.4	32.0	48.0	64.0	96.0	128.0
NB 4 kHz	5.9 VBR	7.2	8.0	9.6	13.2	16.4	24.4					

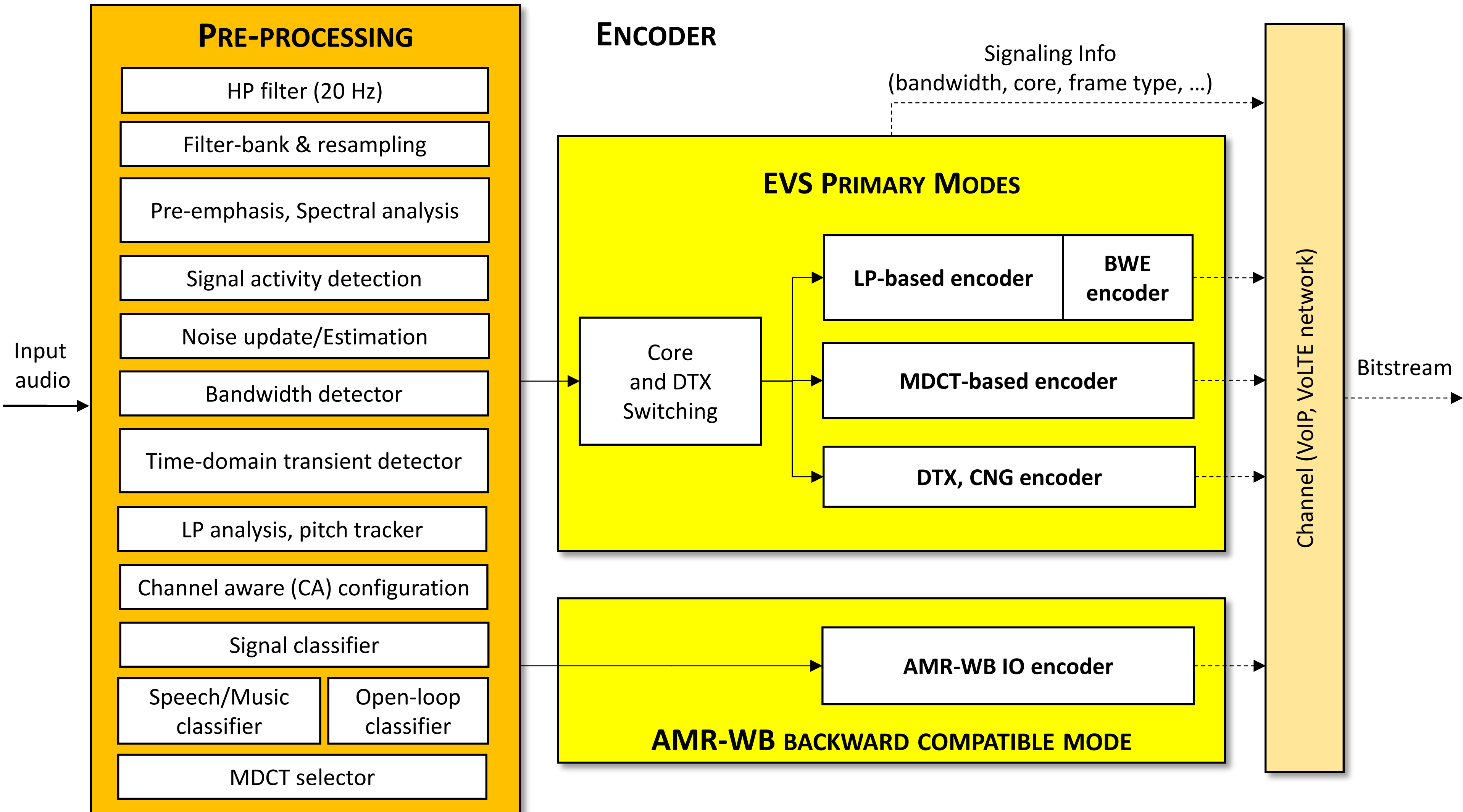
← ACELP/MDCT → ← MDCT →

- Supported sampling-rates: 8 kHz, 16 kHz, 32 kHz, 48 kHz
- Bandwidth detector → automatically switches to effective bandwidth
- Seamless switching between any operating-points → adapt to transmission-channel
- Bitstream compatibility to all AMR-WB modes

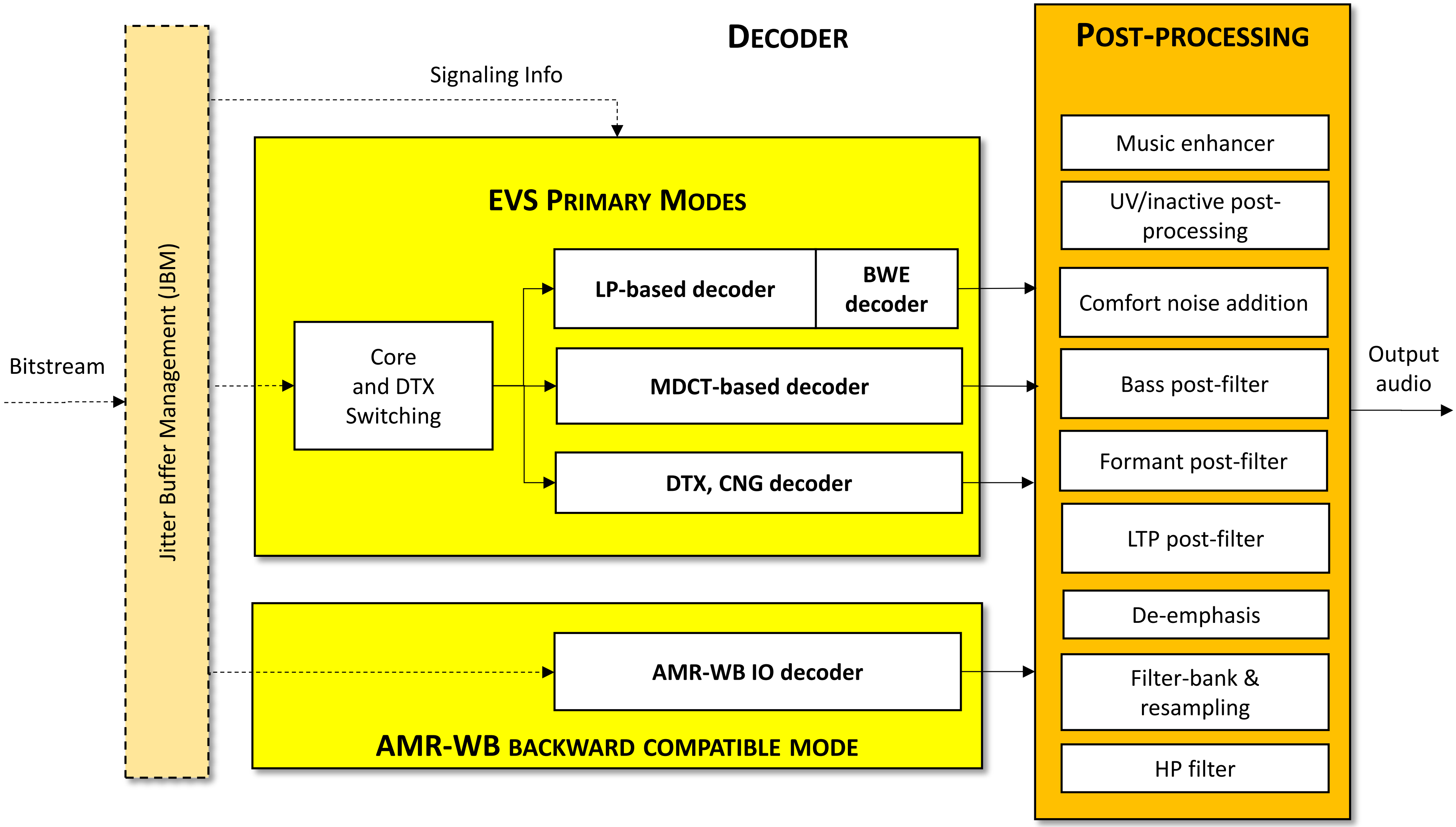
EVS Codec features

Feature	Property
Narrowband (NB) operation	5.9 kbps (VBR), 7.2-24.4 kbps
Wideband (WB) operation	5.9 kbps (VBR), 7.2-128 kbps
	Enhanced interoperation with all AMR-WB modes: 6.6 – 23.85 kbps
Super-Wideband (SWB) operation	9.6-128 kbps
Fullband (FB) operation	16.4-128 kbps
Smart bandwidth control	Optimized bandwidth operation at each rate
VAD/DTX/CNG	Available at all rates, required for 5.9 kbps VBR
Channel-aware mode	Available at 13.2 kbps WB and SWB
Packet-Loss-Concealment	Cutting-edge, included in standard
Jitter buffer management (JBM)	Cutting-edge, included in standard
Rate adaptation support	Seamless rate switching on 20 ms frame basis
Audio sampling rate conversion	Decouples input/output audio sampling rates from codec bandwidth
Algorithmic delay	32 ms

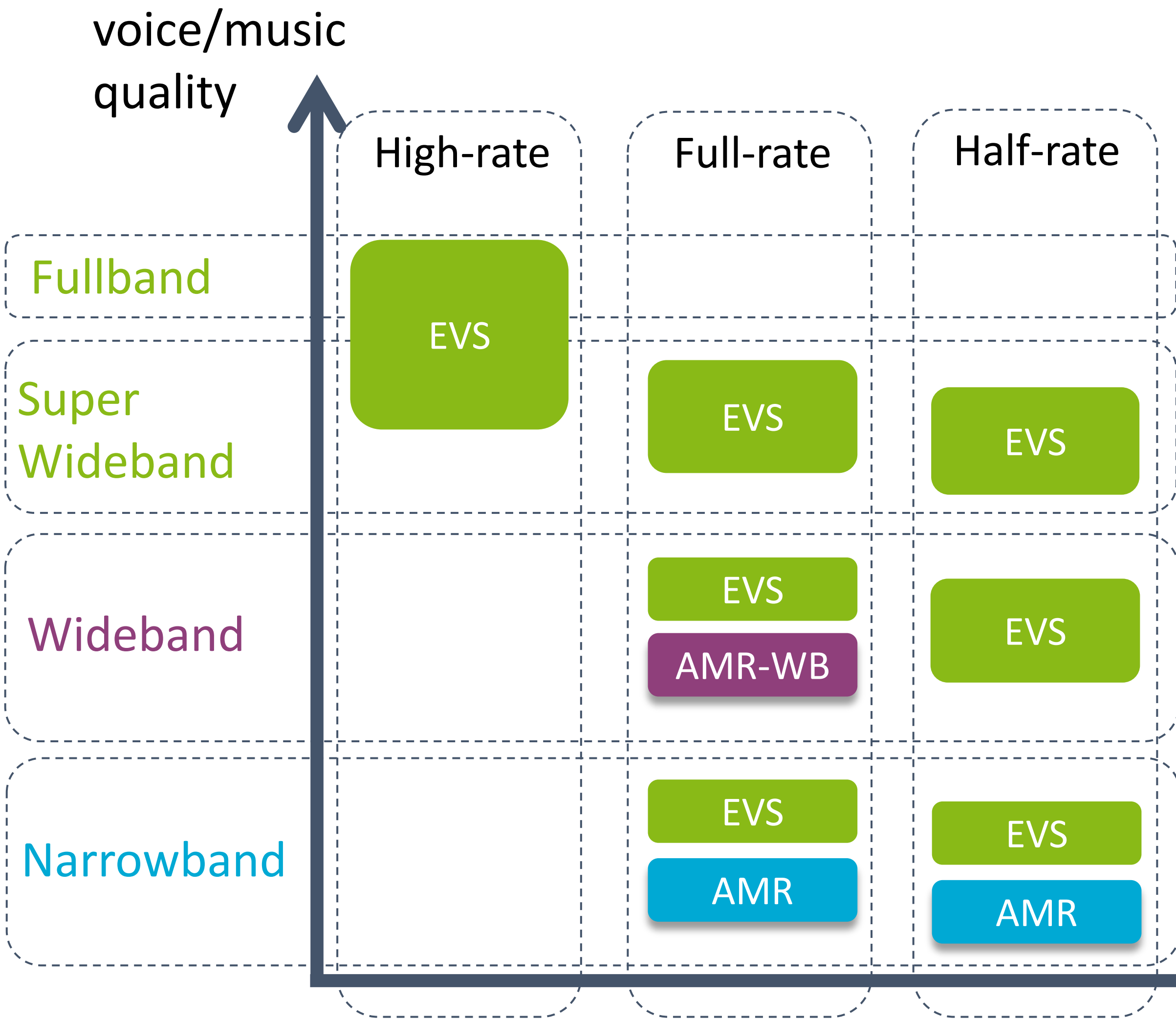
Encoder Block Diagram



Decoder Block Diagram



Benefits of EVS Codec



EVS codec enhances HD voice by:

- Increasing the voice capacity with same or even better quality
- Enhancing the voice and music quality with same capacity
- Can be used for high quality music services, preferably in fullband mode at high rates > 13.2 kbps

Voice demo

	Transport bitrate [kbps]	Narrow band 2.5G/3G voice (AMR)	HD voice (AMR-WB)	EVS
Half-rate	7.2			
Full-rate	13.2	(1	(2	
High-rate	24.4			(3

Original:

(1 increased quality with same capacity
(2 increased capacity
(3 extraordinary quality

Evolved HD voice for superior voice quality

Music demo

	Transport bitrate [kbps]	Narrow band 2.5G/3G voice (AMR)	HD voice (AMR-WB)	EVS
Half-rate	7.2			
Full-rate	13.2	(1	(2	
High-rate	24.4			(3

(1 increased quality with same capacity
(2 increased capacity
(3 extraordinary quality

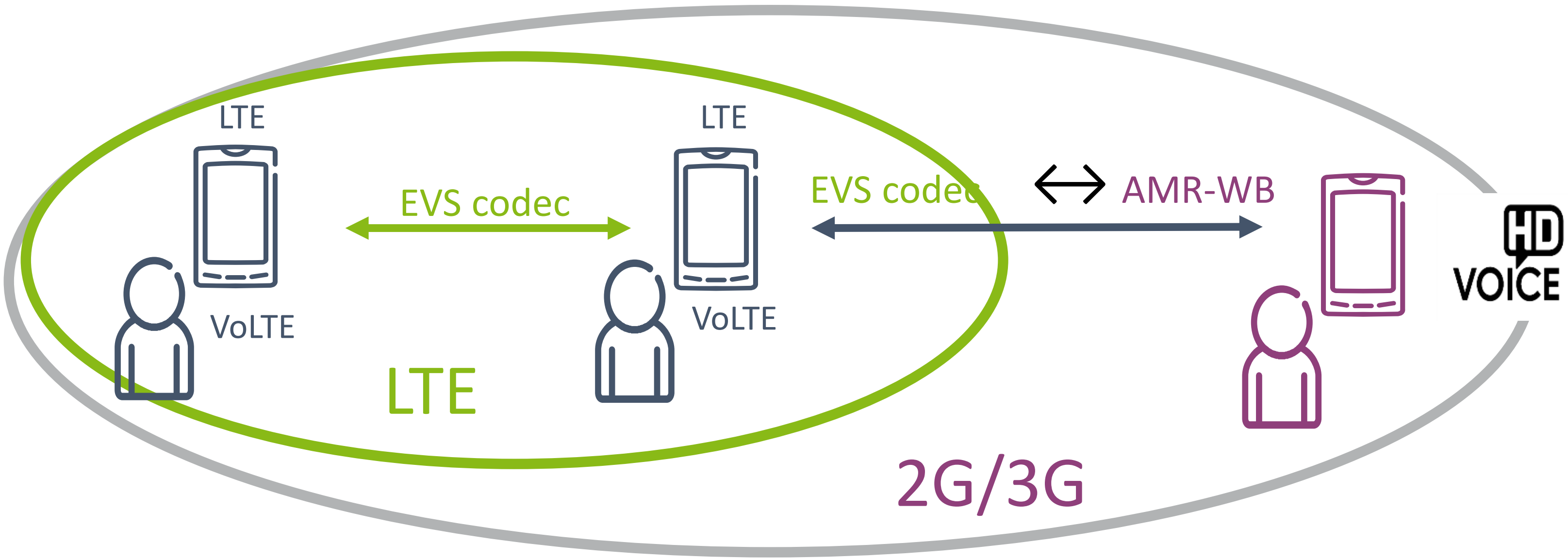
Original:

Excellent music experience in LTE/VoLTE networks

Benefits of EVS

CONSUMER BENEFITS
- Superior voice quality
- Excellent music quality
- Improved experience

OPERATOR BENEFITS
- Globally interoperable – based on 3GPP standards
- Improved telecom grade quality and capacity tradeoff
- Seamless interworking with legacy networks



Excellent voice and music experience

ERICSSON Fraunhofer IIS HUAWEI VoiceAge

EVS Deployments

- Operators
  - Korea:
    - Commercial services since October 2015
  - US:
    - T-Mobile: commercial services launched April 2016
  - Japan:
    - NTT DOCOMO: commercial services launched May 2016
  - Germany:
    - Vodafone: commercial services launched May 2016
  - More operators in various regions interested
- Devices
  - Samsung Galaxy 6 Edge Plus, Galaxy Note 5, Galaxy S7 and S7 edge
  - LG G5
  - Sony Xperia X Performance
  - AQUOS ZETA
  - Disney Mobile on docomo
  - Arrows SV
  - More vendors and models to come …
- Infrastructure
  - Ericsson: Product support in SBG, BGF and MRF since Q1/2016
  - Huawei: Product support since 2016
  - Nokia Networks: Product support since May 2016
- Interoperability Testing (IoT)
  - Bilateral IODT testings between terminal and infrastructure sides since last year
  - IMTC ready to host IOT test on multi-lateral scale

Part 2: Performance, Application Scenarios, Demos

presented by Jon Gibbs, Huawei Technologies Co. Ltd

ERICSSON Fraunhofer IIS HUAWEI VoiceAge www.voiceage.com

Part 2: Outline

- Performance
- Demos
- Application Scenarios

EVS Performance Evaluation Method

- 3GPP in common with ITU-T Recommendations uses naive listeners and ITU-T P.800 tests
  - Representative of the (mobile) telephony customer base
  - Clean speech, speech in background noise and music/mixed content categories in error-free & IP packet/frame loss
  - Independent (of the Developers) Host (Processing) Lab, Listening Labs and Global Analysis Labs used throughout.
  - 42 Experiments conducted between Selection (24) and Characterization (18) costing approximately €1.1M
  - Each major experiment conducted in two different languages (different language group) in different listening laboratories
    - 48 P.800 Tests during Selection
    - 24 P.800 Tests during Characterization
  - Results Documented in 3GPP TR 26.952
    - 10 different languages employed
    - Assessment of potential language dependence – nothing indicated
    - Only 2 out of 389 Requirements failed systematically (0.5%)
    - Only 38 out of 295 Objectives failed systematically (13%)

Performance Summary

Current
Narrowband (NB)
Speech Service
(AMR)
4.75 kbps – 12.2 kbps

Current
Wideband (WB)
HD Voice Service
(AMR-WB)
6.6 kbps – 23.85 kbps

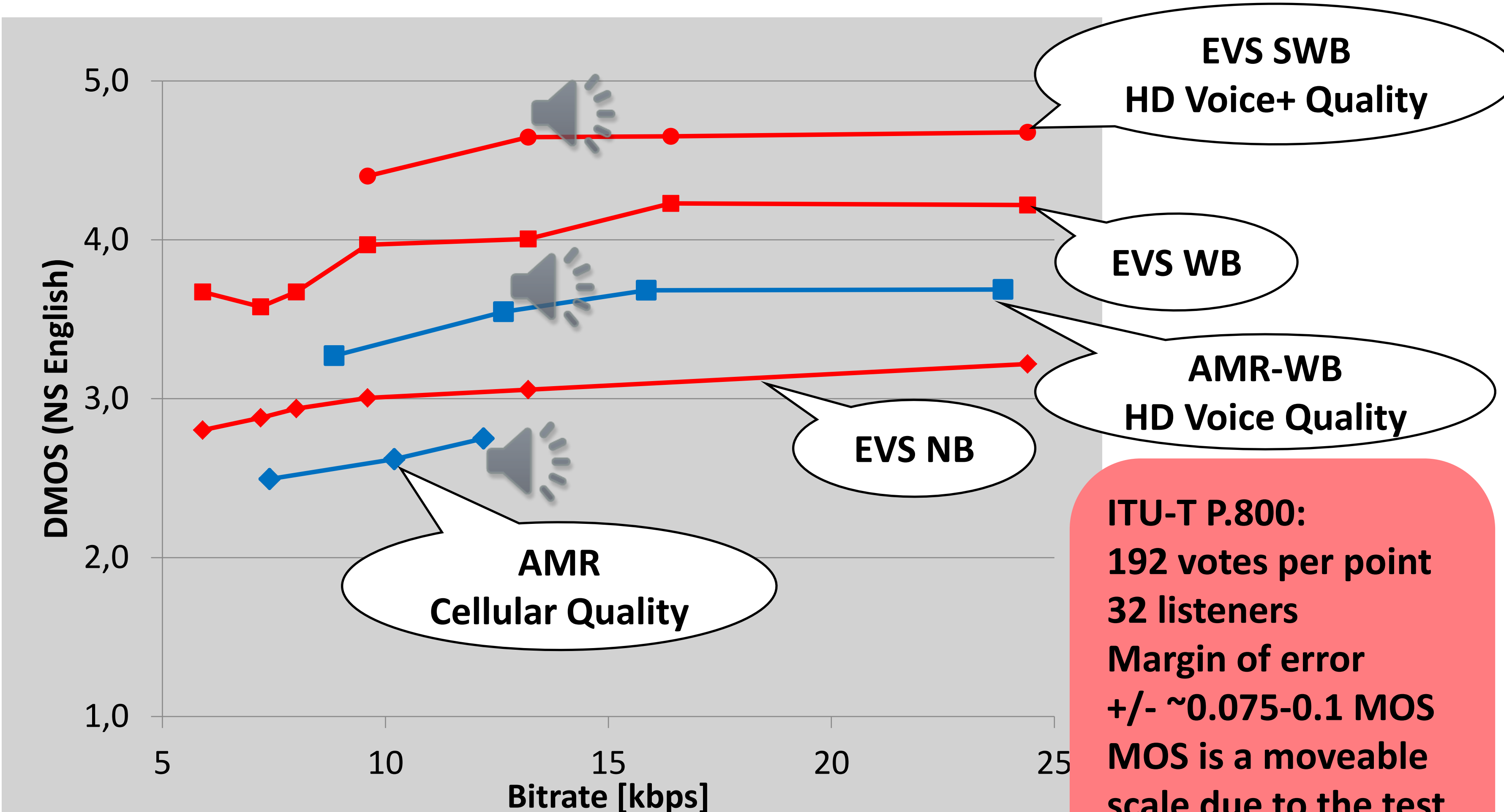
EVS - Enhanced Voice Services

- Same quality available at Lower bitrate
  5.9 kbps – 24.4 kbps
- Higher quality for the same bitrate
- Higher quality in IP packet/frame loss
- Higher Music and mixed content quality

- Same quality available at Lower bitrate
  5.9 kbps – 128 kbps
- Higher quality for the same bitrate
- Higher quality in IP packet/frame loss
- Higher Music and mixed content quality

- **New Super-Wideband (SWB) (9.6 kbps – 128 kbps) and Fullband (FB) (16 kbps – 128 kbps) HD Voice+ Service**
- **HD Voice+ Service consistently better than NB and HD Voice (WB) Service**

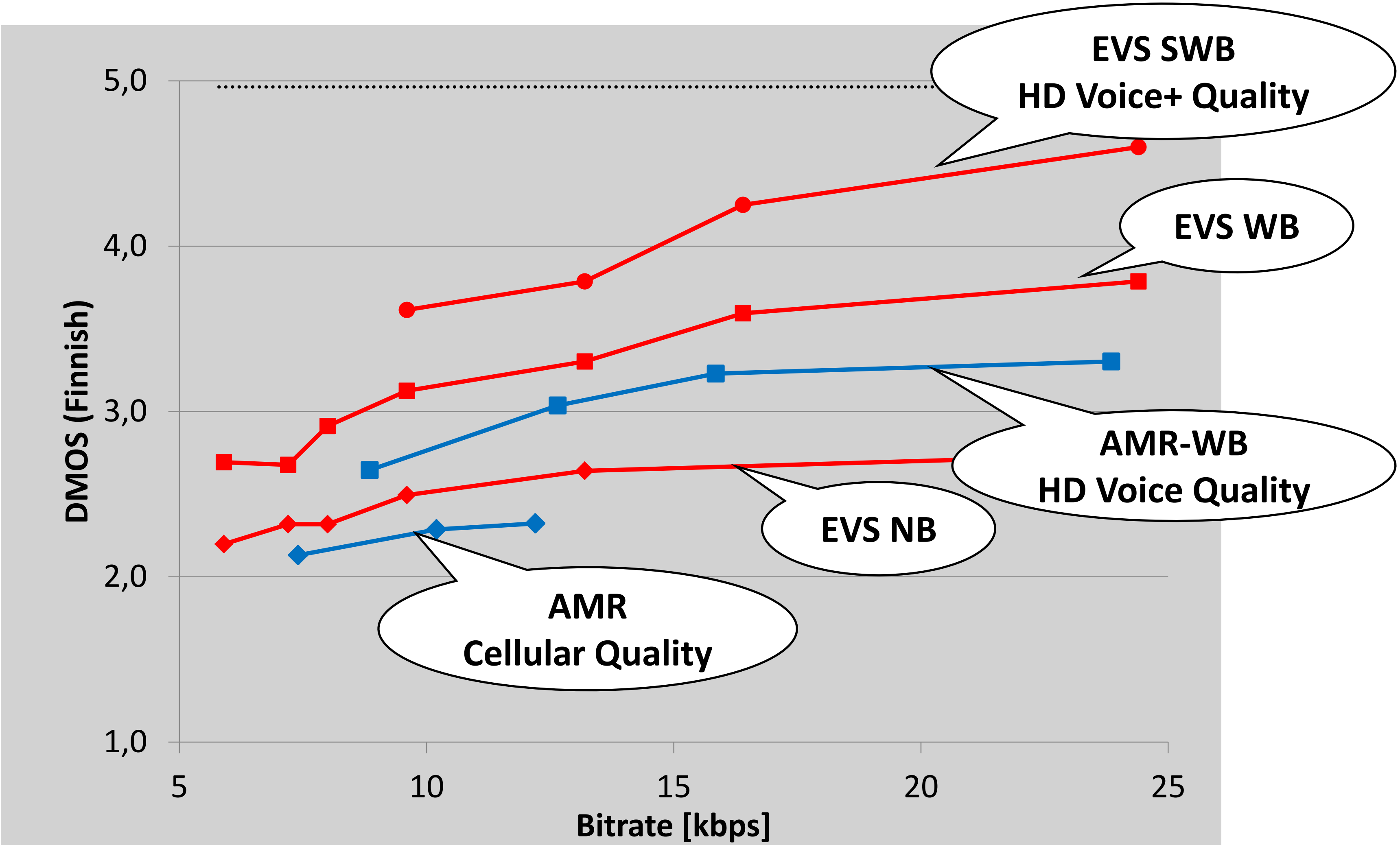
EVS Clean Speech Quality



ITU-T P.800:
192 votes per point
32 listeners
Margin of error
+/- ~0.075-0.1 MOS
MOS is a moveable scale due to the test context.

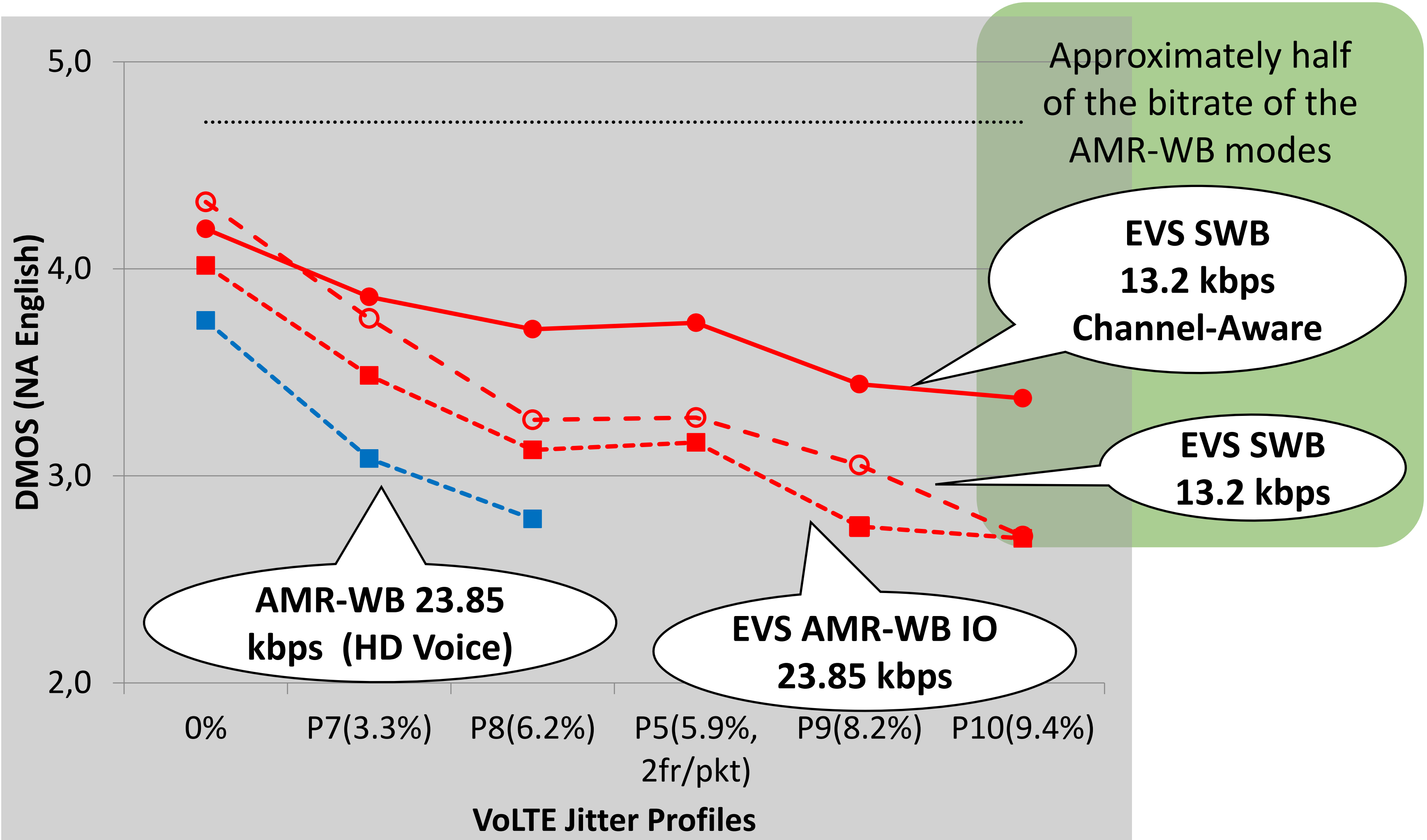
ERICSSON Fraunhofer IIS HUAWEI VoiceAge

EVS Noisy Speech Quality (Car Noise 20dB SNR)



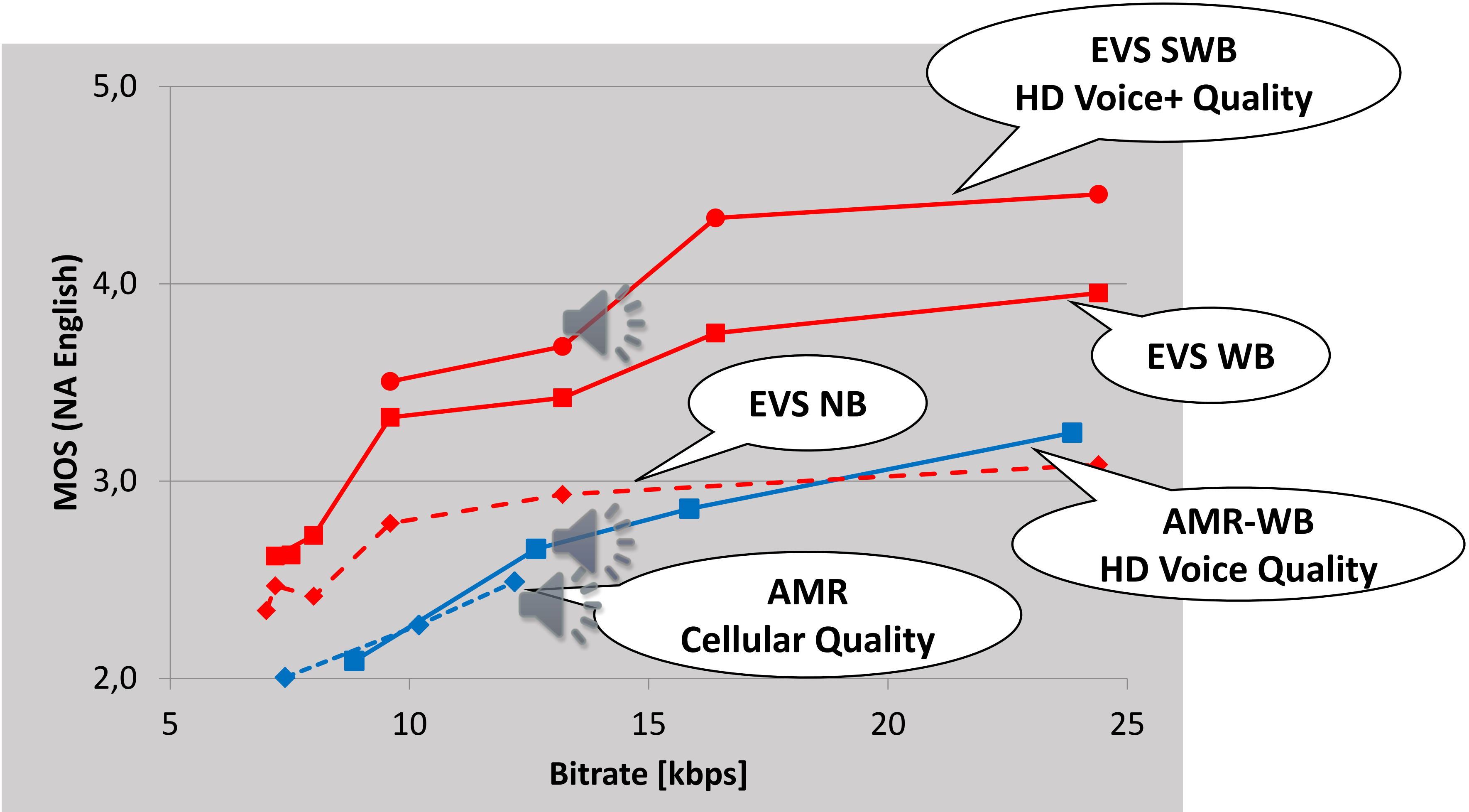
ERICSSON Fraunhofer IIS HUAWEI VoiceAge

EVS SWB - Clean Speech (VoLTE Jitter)

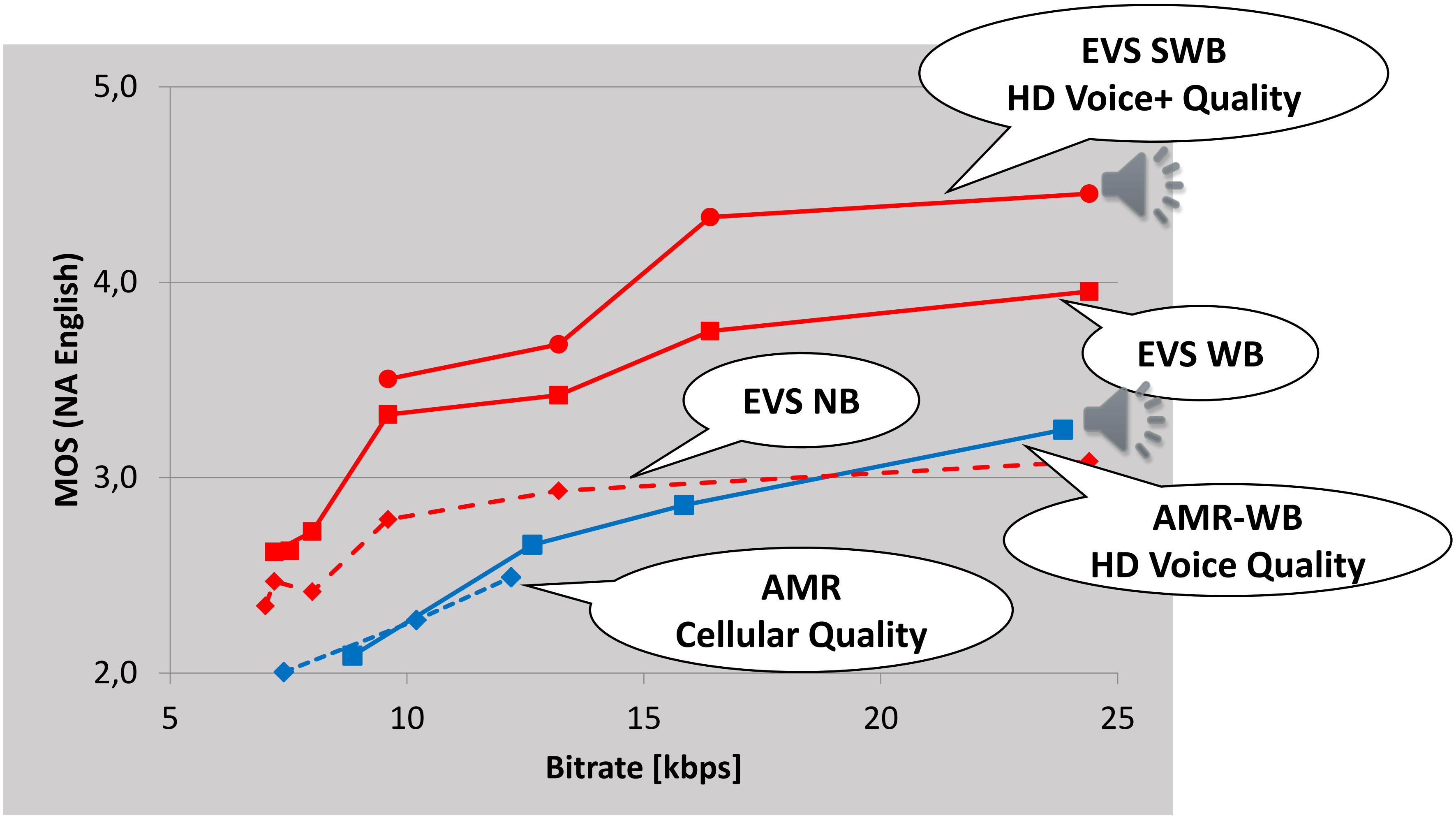


ERICSSON Fraunhofer IIS HUAWEI VoiceAge

Music & Mixed Content

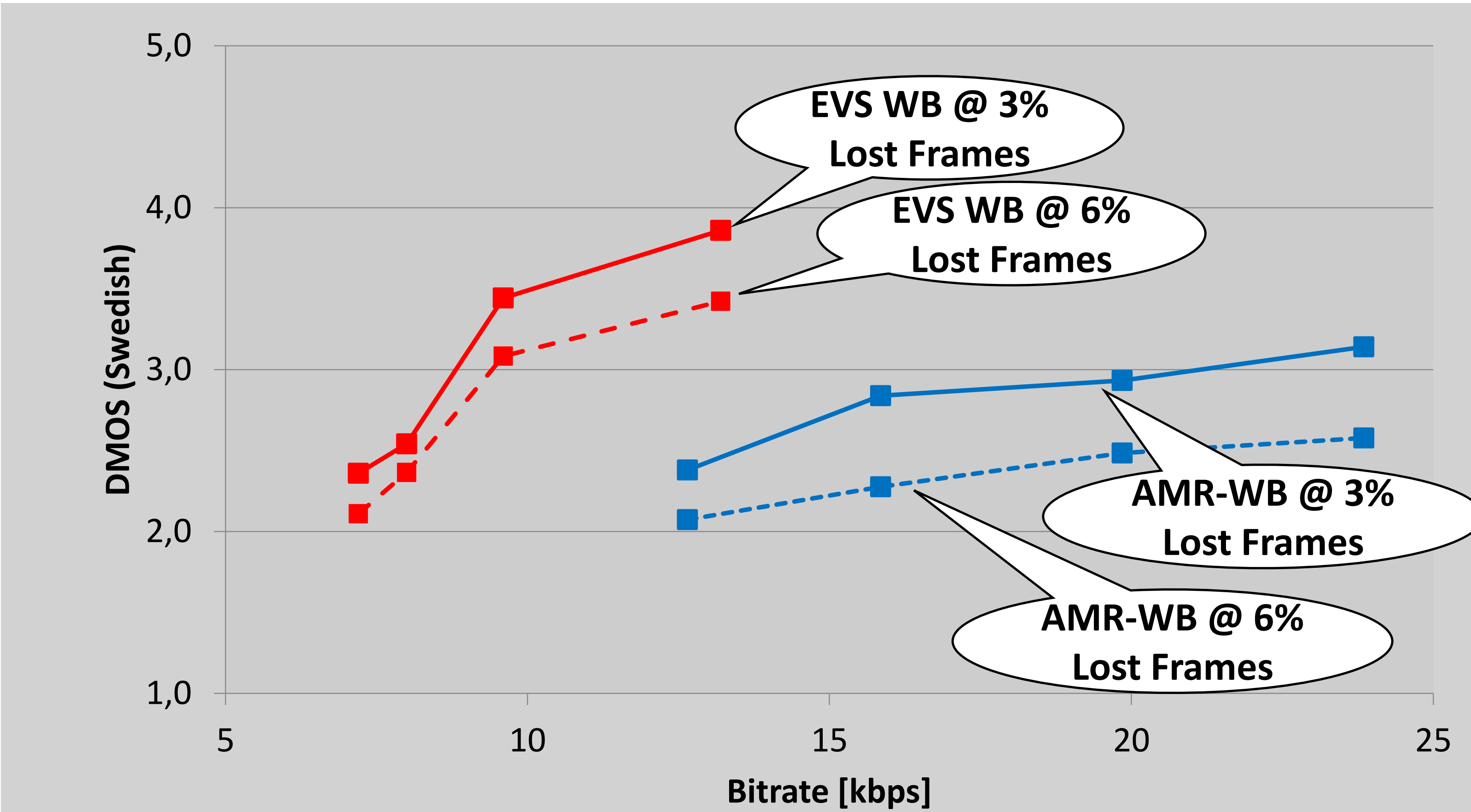


Music & Mixed Content



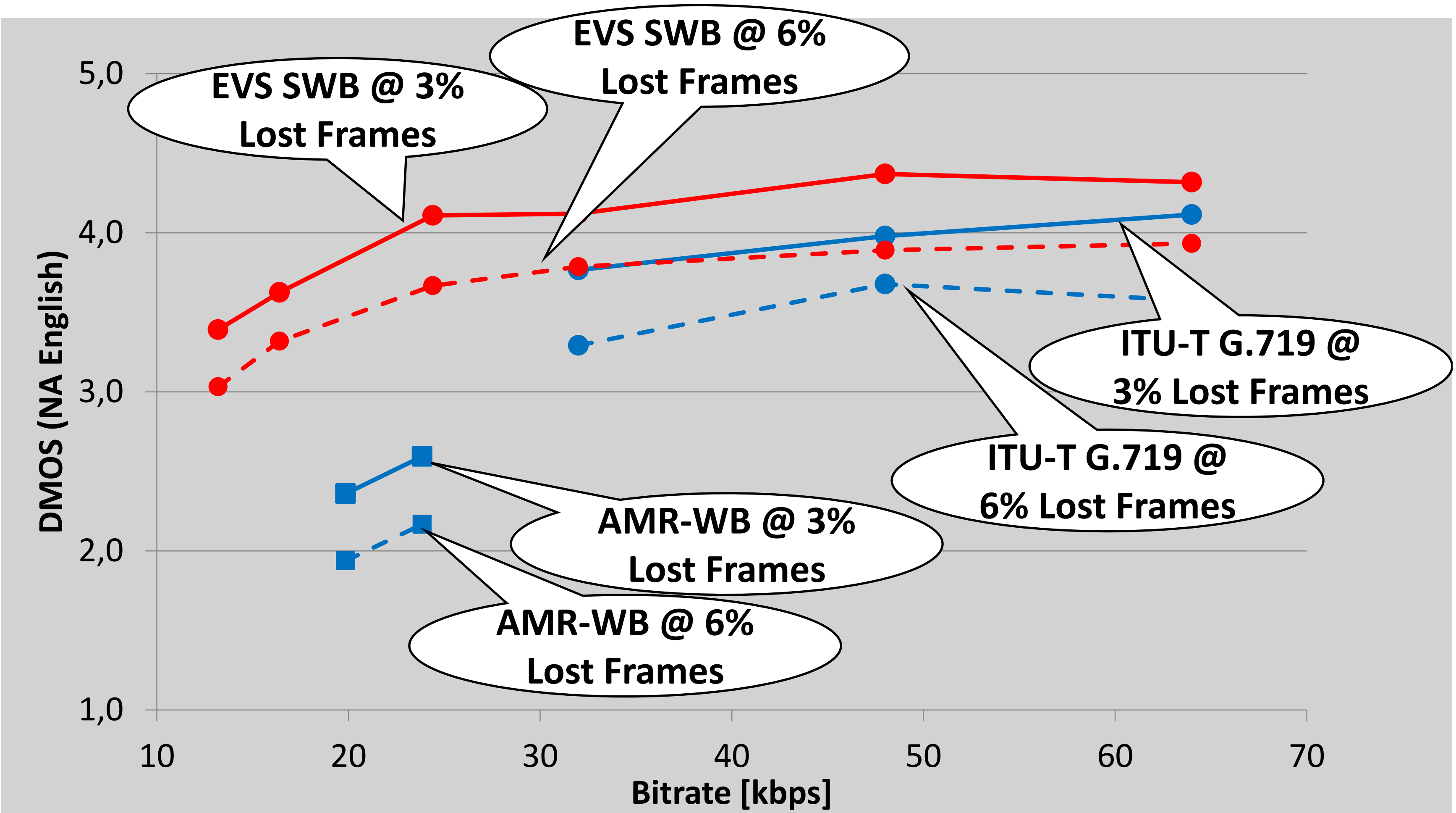
ERICSSON Fraunhofer IIS HUAWEI VoiceAge

WB Music & Mixed Content (Frame Losses)



ERICSSON Fraunhofer IIS HUAWEI VoiceAge

SWB Music & Mixed Content (Frame Losses)



ERICSSON Fraunhofer IIS HUAWEI VoiceAge

Capacity Enhancement – EVS at ½ Bit-rate

- **SWB EVS at 13.2 kbps vs AMR-WB at 23.85 kbps**

Original

AMR-WB (23.85 kbps)

EVS (SWB 13.2 kbps)

- **SWB EVS at 13.2 kbps in FERs vs AMR-WB at 23.85 kbps**

Original

AMR-WB (23.85 kbps)

EVS (SWB 13.2 kbps)

Application Scenarios

- Mobile Telephony & Audio Conferencing
  - NB (narrowband PSTN interconnection)
  - WB (HD Voice) – Optional for GSMA HD Voice
  - SWB (HD Voice+) – Mandatory Codec for GSMA HD Voice+
  - Handset, headset and handsfree
  - High Packet Loss & Delay Jitter Resilience
- VoLTE and VoWiFi/VoIP
- In-call music and music-on-hold
- Mission Critical Push to Talk
  - EVS (SWB) is an optional codec
  - Very conservative industry
  - EVS able to demonstrate intelligibility and coverage gains over AMR-WB

Part 3:
Coding of Speech in EVS

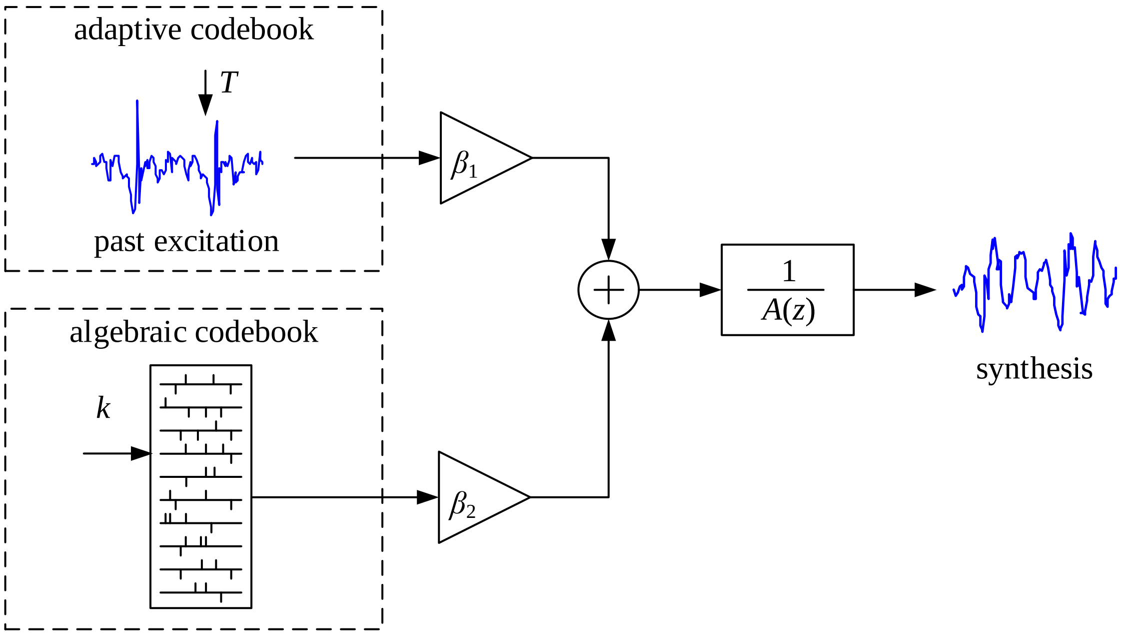
presented by Václav Eksler, VoiceAge Corporation

Part 3: Outline

- Introduction
- Improved variant of ACELP
- Extended classification of input signal
- Post-processing enhancements
- Coding of upper band
- Advanced error resilience
- Source Controlled Variable Bitrate Coding (SC-VBR)
- AMR-WB backward compatibility and improvements
- Discontinuous Transmission and Comfort Noise Generation

Introduction

- Most of the current speech codecs, e.g. AMR-WB, are based on code-excited linear prediction (CELP) model
- Algebraic CELP (ACELP) employs large codebooks of fixed pulses



Improved ACELP in EVS 1/2

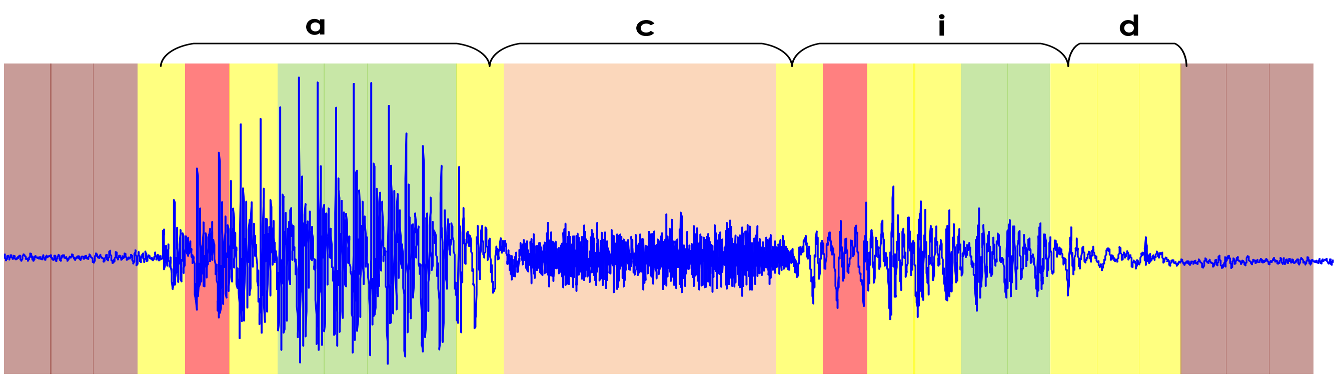
- Key design points:
  - higher bandwidth → ACELP in the lower band, BWE in the higher band
  - advanced error resilience
  - much more general content
- Support of 12.8 kHz and 16 kHz internal ACELP sampling rate
- Significant **improvements** through extended signal classification
  - active/inactive, bandwidth, speech/music/mixed, clean/noisy, …
- **Generic Signal Coding** (GSC)
  - LP-based time-frequency mode
- **Transition Coding**
  - extended to encode strong onsets
  - employed in switching frames

Improved ACELP in EVS 2/2

- **Unvoiced Coding**
  - excitation composed of Gaussian noise combined with algebraic codebook
- Frequency-domain component of the excitation at higher bitrates
- **Bandwidth Extensions** (BWEs)
  - Time domain BWE for active speech
  - Frequency domain BWE for inactive speech and music/mixed segments
- **Optimizations** to
  - Voice Activity Detection (VAD)
  - Open-loop pitch search
  - Adaptive lag-windowing
  - Quantization and indexing of LP coefficients
  - etc.

Speech signal classes

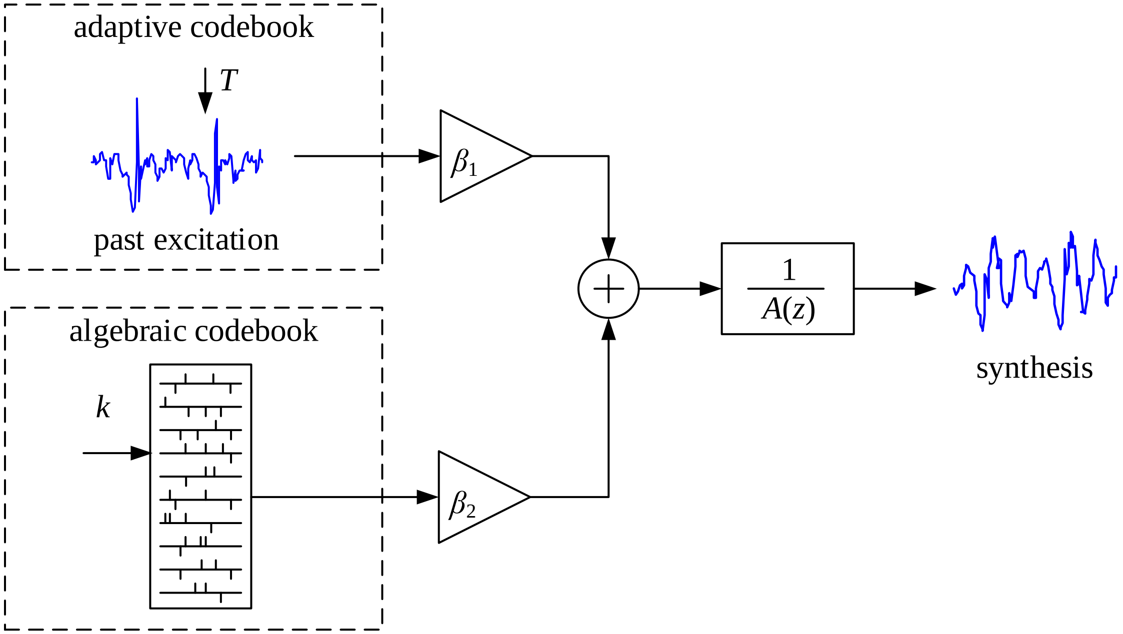
- Significant improvements through detailed determination of speech signal classes in the preprocessing
  - Inactive – speech or audio activity not detected
  - Unvoiced – unvoiced speech frames
  - Voiced – quasi-periodic stable active segments
  - Transition – improve robustness and encode strong onsets
  - Generic – all other speech frames
- Example: classification of word "*acid*"

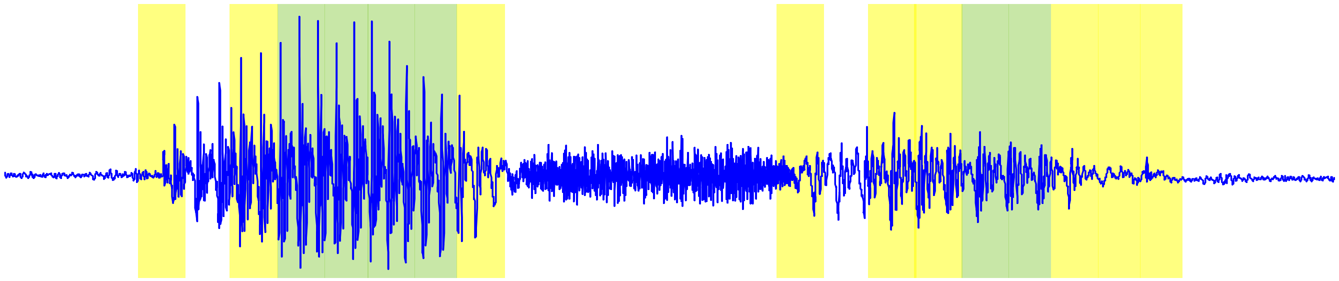


ERICSSON Fraunhofer IIS HUAWEI VoiceAge www.voiceage.com

Generic Coding – lower bitrates

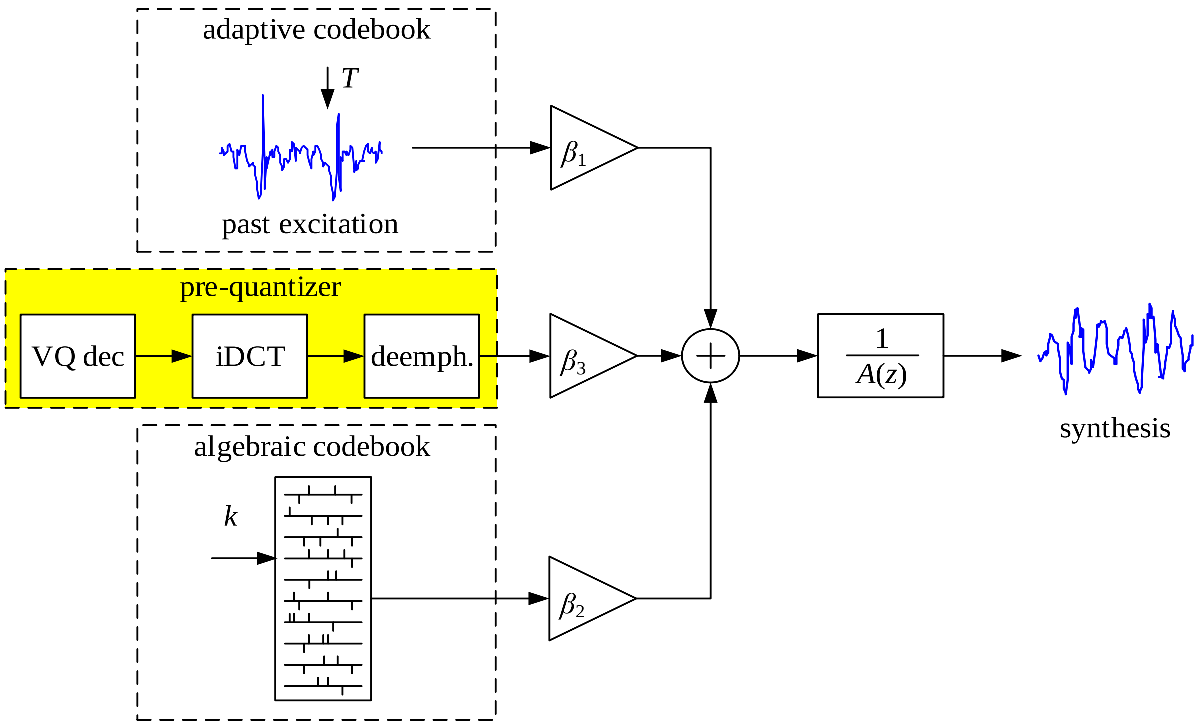
- Traditional adaptive codebook and fixed algebraic codebook
- Voiced Coding – higher bit-budget to algebraic codebook





Generic Coding – higher bitrates

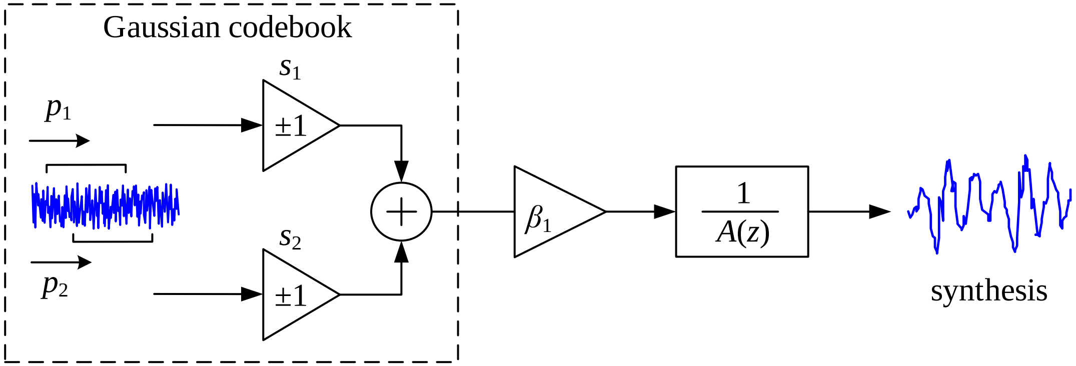
- Overcomes exploding complexity problem when searching for all possible algebraic codebook vectors (e.g. at 32 kbps: 106 bits, i.e. ~ $8x10^{31}$ vectors)
- → Frequency-domain component of the excitation

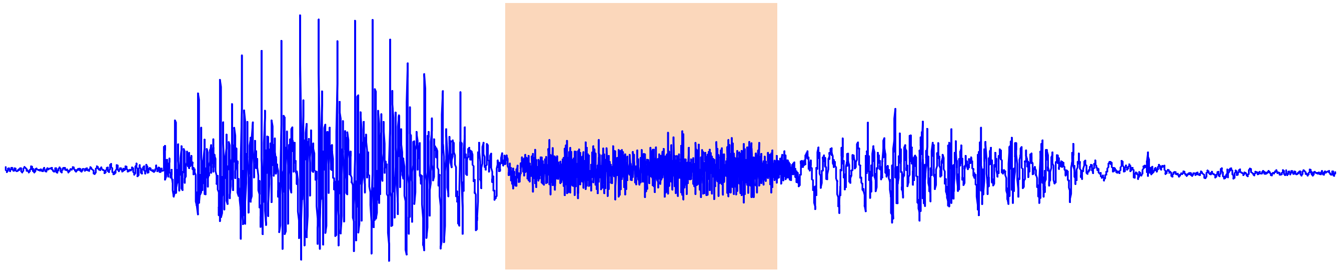


ERICSSON Fraunhofer IIS HUAWEI VoiceAge www.voiceage.com

Unvoiced Coding – lower bitrates

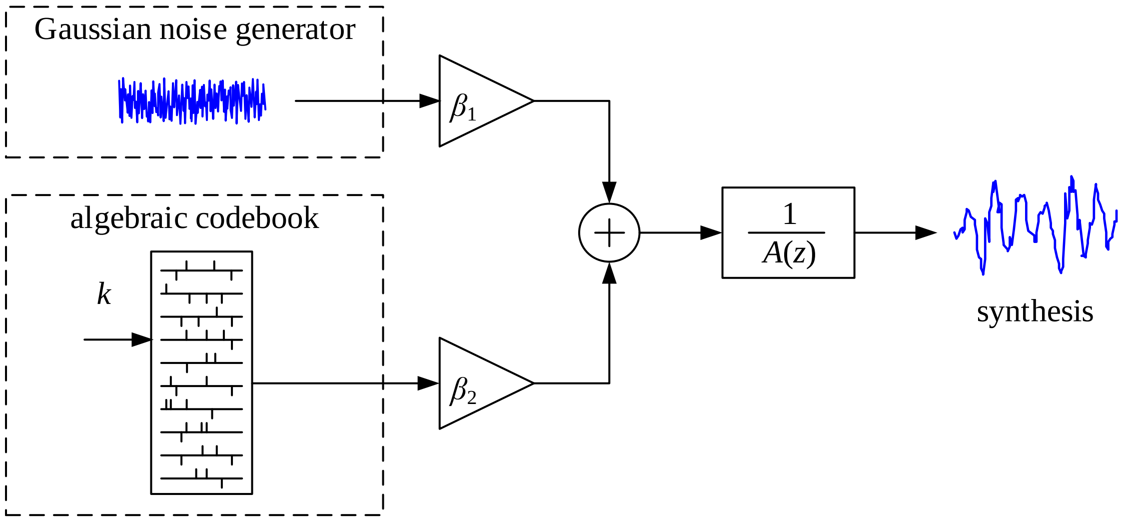
- Excitation composed of two vectors selected from a linear Gaussian codebook

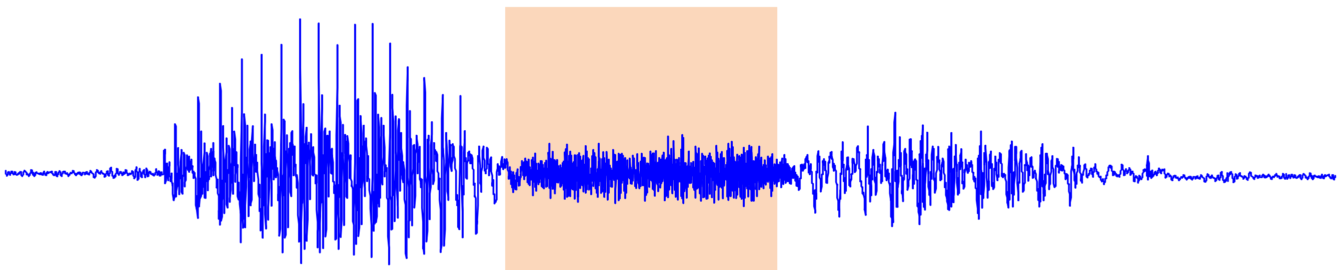




Unvoiced Coding – higher bitrates

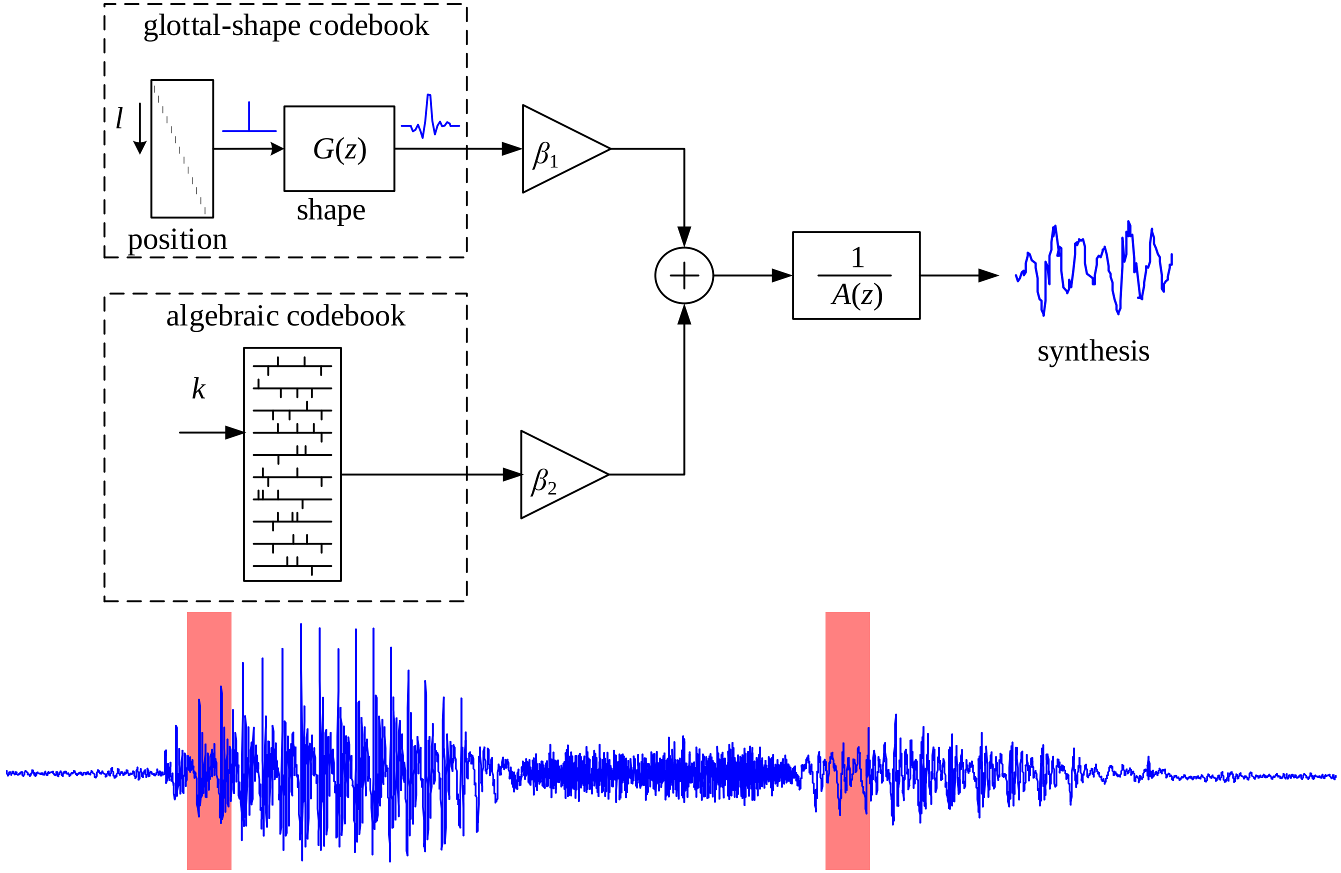
- Excitation composed of Gaussian noise combined with algebraic codebook





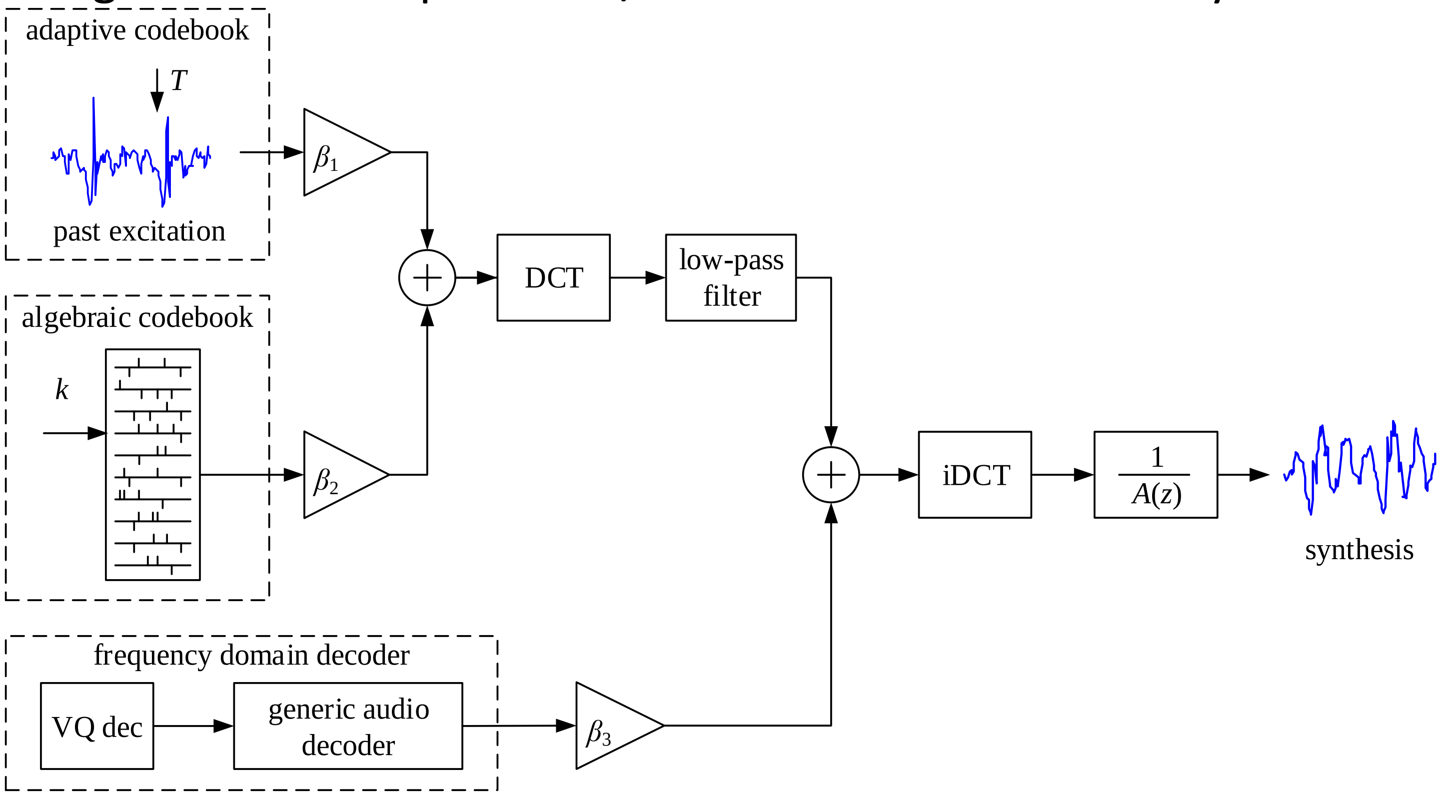
Transition Coding

- Adaptive codebook replaced by codebook of glottal shapes → significantly limits the usage of past information
- Protects frames after onsets; encodes strong onsets and switching frames



Generic Signal Coding

- New coding mode for efficient coding of generic audio signals, particularly music, at low bitrates
- Combines encoding of excitation in time and frequency domain
- Rectangular windows prior DCT/iDCT → no additional delay

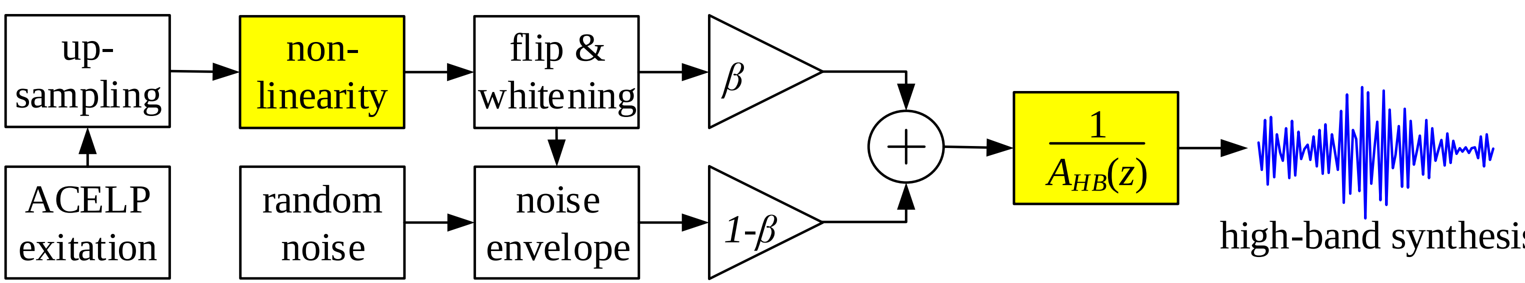


Decoder post-processing

- **Bass post-filter**
  - Improved low-frequency pitch enhancement
  - Controlled by signal classification
- **Formant post-filter**
  - Formant sharpening
  - Controlled by signal classification
- **Music post-filter**
  - New low-delay technique to enhance music at lowest bitrates
- **Comfort Noise Addition for noisy speech**
  - New technique to improve rendering of background noise at lower bitrates
  - Artificial noise injected in both active and inactive segments
  - → Masks coding artifacts and discontinuities
  - → Compensates the loss of energy in the background noise

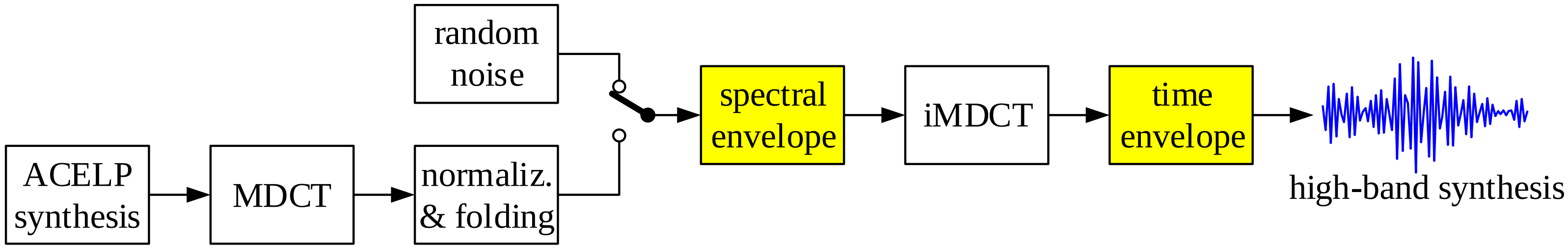
Time domain BWE

- New, highly efficient BWE on top of ACELP
- Extends the decoded spectrum to WB (up to 8 kHz), SWB (up to 16kHz), or FB (up to 20 kHz)
- BWE bitrate of 0.30 kbps – 3.0 kbps
- Time-domain envelope
  - **separate LPC model** in the high band
- High-band excitation signal
  - derived from low-band excitation signal using a **non-linear harmonic modelling**
  - adaptive whitening
  - noise modulation and mixing



Frequency domain BWE

- A novel multi-mode frequency domain BWE with relaxed synchronization on top of GSC
- Blind (0 kbps) or guided with bitrate of 0.30 kbps – 3.0 kbps
- **4 modes**: Transient, Harmonic, Normal, Noise
- A combination of adaptive **spectral envelope** and **time envelope** coding, derived from the high-band input signal
- High-band excitation signal generated by
  - normalizing the selected region of the low band with an adaptive normalization length, or
  - random noise
- Only a low algorithmic delay available → **relaxed time alignment** between the high-band excitation and its envelope

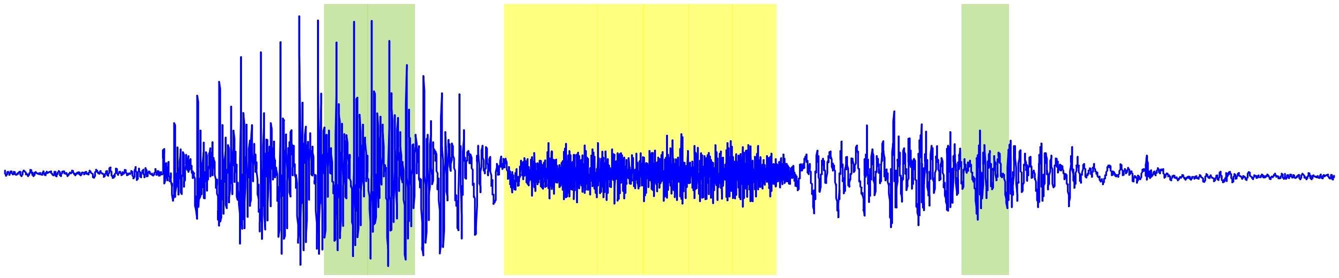


Advanced Error Resilience

- Multiple innovative **highly robust** measures to provide error resilience to packet losses in mobile systems
  - Minimization of inter-frame dependencies
    - avoid error-propagation
    - fast recovery after lost packets
  - Improvements to various blind or guided concealment techniques
    - improved pitch extrapolation, improved pulse resynchronization, guided LP filter concealment, etc.
    - focus both on concealed frame(s) and recovery frames
    - both single errors and long burst of errors
- Built-in **Jitter Buffer Management**
  - Compensation for transmission delay jitter (late packets)
- **Channel-Aware Coding** at 13.2 kbps
  - Partial redundancy transmitted in later frames
  - The side-info transmission is source/channel controlled

Source Controlled VBR

- Targets an average bitrate of 5.9 kbps for active speech
- Source controlled switching between 2.8 kbps, 7.2 kbps and 8.0 kbps frames
- Efficient coding modes for 2.8 kbps frames:
  - prototype pitch period (PPP) → stationary voiced frames
    - Pitch cycles in these frames are stable
    - Transmit just one representative prototype pitch period
    - Derive remaining pitch cycles by interpolation
  - noise-excited linear prediction (NELP) → unvoiced frames



AMR-WB Backward Compatibility

- EVS offers AMR-WB **interoperable mode** (AMR-WB IO)
  - Full bitstream compatibility for all AMR-WB bitrates
- AMR-WB IO offers **improvements** over legacy AMR-WB
  - Improved error concealment
  - Better quality through EVS post-processing modules
    - Bass Post-Filter, Comfort Noise Addition, Formant Post-Filter
  - Better music quality through music enhancer
    - DCT based suppression of quantization noise
  - Better noisy speech quality through unvoiced/inactive post-processing
    - Smooth synthesis output by modifying the excitation in DCT domain
  - Better presence through a higher audio bandwidth
    - New bandwidth extension up to 7.8 kHz
  - Fixed-point code: Better reproduction of low-level input signals through dynamic scaling

Discontinuous Transmission

- Improved Discontinuous Transmission (DTX) for efficient use of spectrum and battery life in mobile communication
- Background noise replaced by **Comfort Noise Generation** (CNG) at the decoder
- **Silence Insertion Description** (SID)
  - low-rate parametric representation of the noise (2.4 kbps)
  - sent no more than once in every 8 frames (160 ms)
- Enhanced versatility
  - improved LP-based CNG
  - new frequency domain based CNG
  - CNG type selected based on the background noise characteristics

Part 4:
Coding of Mixed/Music Content in EVS

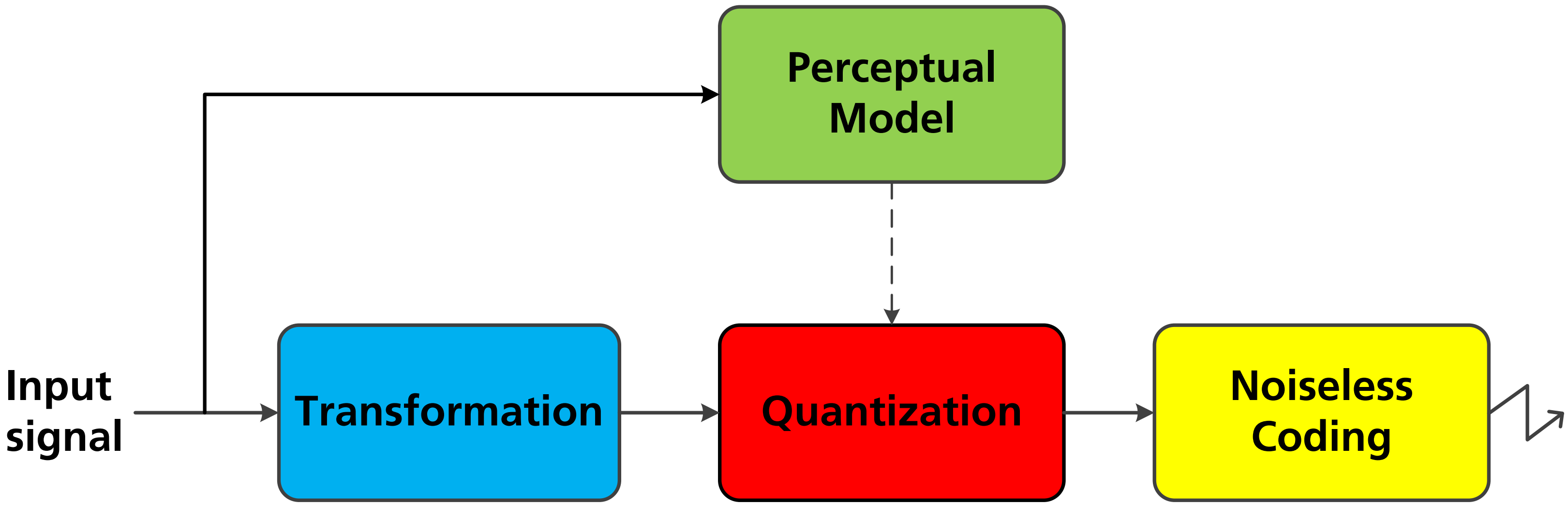
presented by Guillaume Fuchs, Fraunhofer IIS

Part 4: Outline

- Introduction
- System constraints
- System overview
- Envelope coding & noise shaping
- Windowing & switching
- Optimized spectral coding
- Noise and gap filling
- Concealment
- Post-processing

Introduction

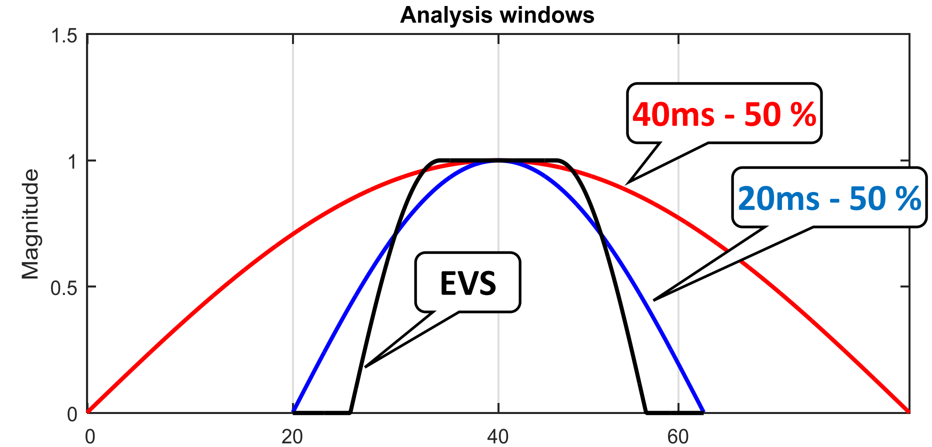
- Most of modern generic audio coders are built over a Modified Discrete Cosinus Transform (MDCT)
- **Redundancy** is exploited by both MDCT and noiseless coding
- Quantization and parametric coding are perceptually motivated and exploit the **irrelevancy** in the signal



System constraints: low delay

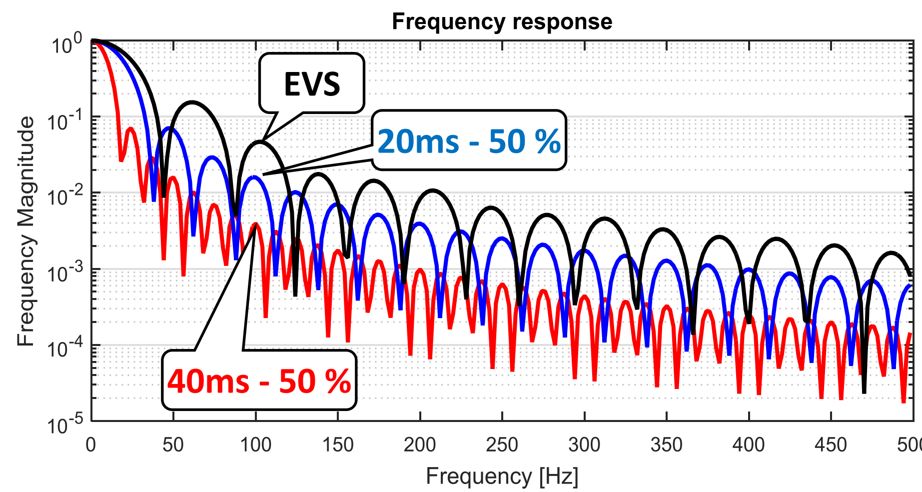
32ms delay coding system:

- Framing: **20ms**
- Overlap: 8.75ms (**21.88%**)
- Additional delay for other components: 3.25ms



Worse frequency responses than conventional audio coder:

- Lower frequency selectivity
- Higher frequency leakage
- Less efficient especially for tonal items



→ New coding tools are introduced for handling tonal music

ca,'XGrid','on')

ERICSSON Fraunhofer IIS HUAWEI VoiceAge

System constraints: adaptive switching

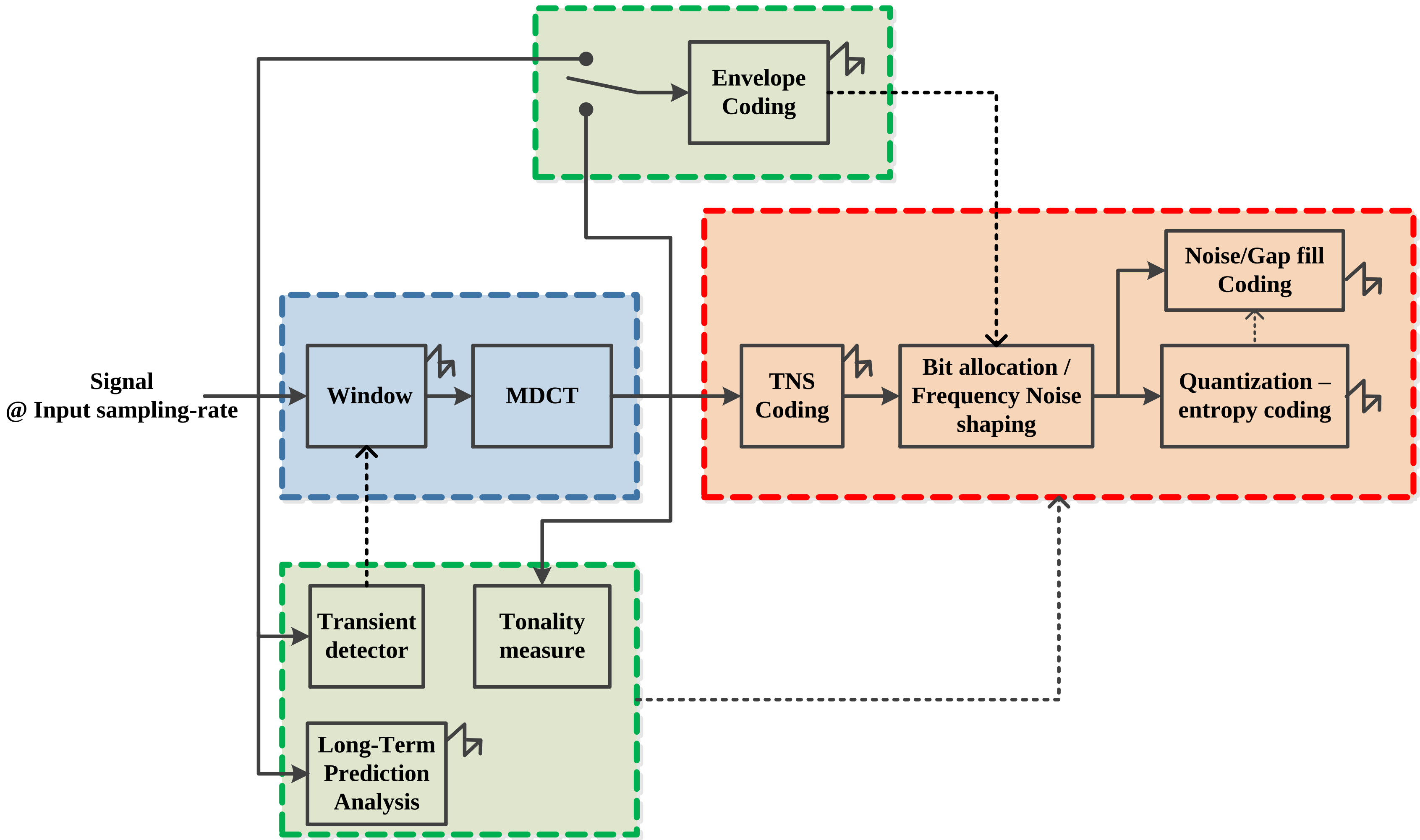
- In EVS, the input signal is classified every 20ms by
  - Speech/music classifier
  - Transient detector
- Depending of the classification a **seamless** and **delayless** switching to a different coding mode or to a different MDCT window is possible:
  - **For speech**
    - Switch to a Time Domain speech coder (ACELP)
  - **For transients**
    - Switch between different time/frequency resolutions
    - Switch between different window shapes
    - Switch to a Time Domain speech coder (ACELP)
- Moreover depending on the different **network conditions** the MDCT coder is able to change on the fly its bit-rate and coded bandwidth.

System constraints: operating points

MDCT-based coding is used for

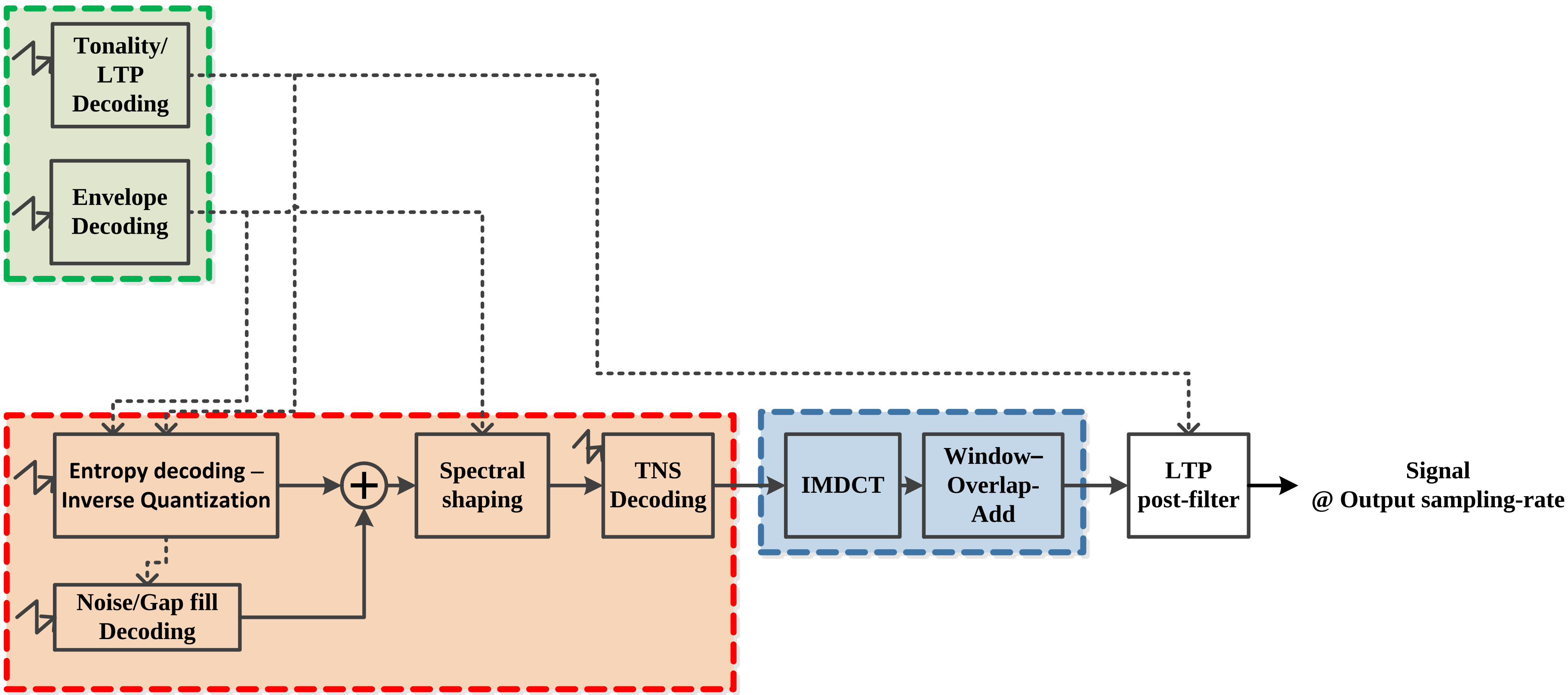
- Different bandwidths and bitrates
  - **Narrowband**: 7.2 – 24.4 kbps
  - **Wideband**: 9.6 – 128 kbps
  - **Super Wideband**: 9.6 – 128 kbps
  - **Fullband**: 16.4 – 128 kbps
- Different signals
  - **Music**: from 7.2 kbps
  - **Background noise**: from 9.6 kbps
  - **Speech**: from 48 kbps

System overview: encoder



ERICSSON Fraunhofer IIS HUAWEI VoiceAge

System overview: decoder



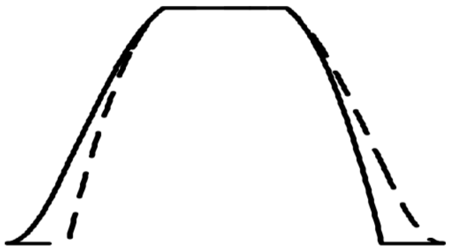
Envelope coding & spectral noise shaping

Two methods are available in EVS for coding the spectral envelope and shaping the quantization noise

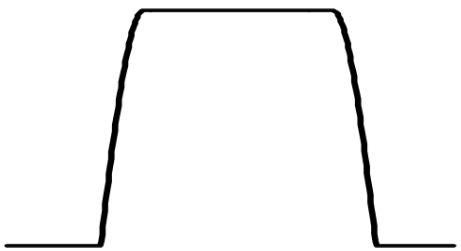
- **Linear Predictive Coding**
  - Compact representation of the spectral envelope (VQ of LSFs)
  - Generic perceptual model derived from weighted LPC coefficients
  - Ease the switching to ACELP
  - Improved and low-delay version of TCX principle from MPEG-USAC
- **Energies of the envelope**
  - Coding of band energies
  - Efficient adaptive Huffman coding of energy differences
  - Bit allocation and noise shaping independent from LPC
  - Improved version of G.719

Windowing & switching

- Asymmetric window: **stationary signals**
  - Lookahead: 8.75ms
  - Overlap: 14.375ms
  - Better frequency response



- Symmetric windows: **transient signals**
  - Smaller overlaps: 3.75 and 1.25ms
  - No time modulation
  - Limit time smearing

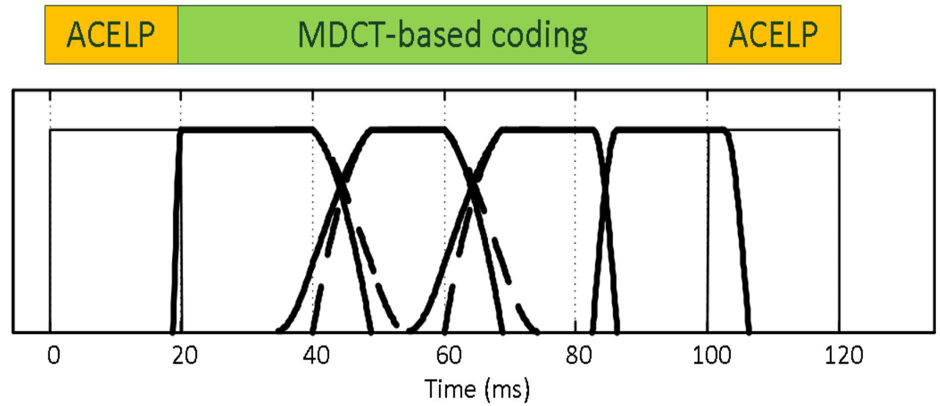


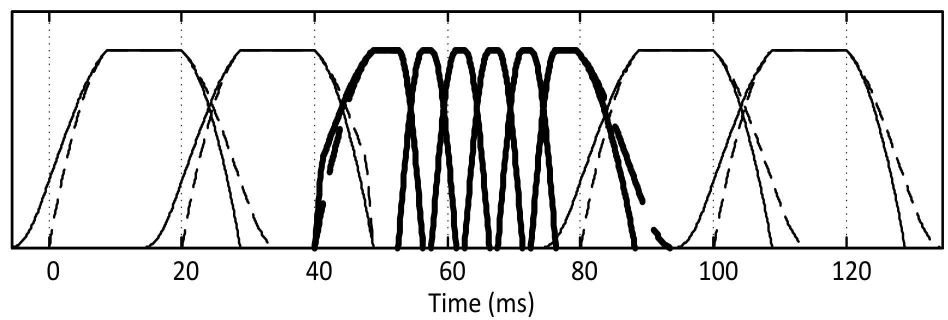
- Short windows: **attacks**
  - 5 and 10ms blocks
  - Better time resolution

ERICSSON Fraunhofer IIS HUAWEI VoiceAge www.voiceage.com

Windowing & switching

- ACELP to MDCT-based coding
  - Enlarged MDCT window, smoothing by ZIR (LPC envelope)
  - Extra ACELP subframe (NRG envelope)
- MDCT-based coding to ACELP
  - Discard overlapping part of MDCT
  - Updating ACELP memories (LPC envelope)
  - First ACELP in Transition Coding (NRG envelope)
- Block switching
  - Delayless switching to 5 or 10ms windows
  - With transition windows
  - Or without transition windows in TDA domain





Optimized spectral coding

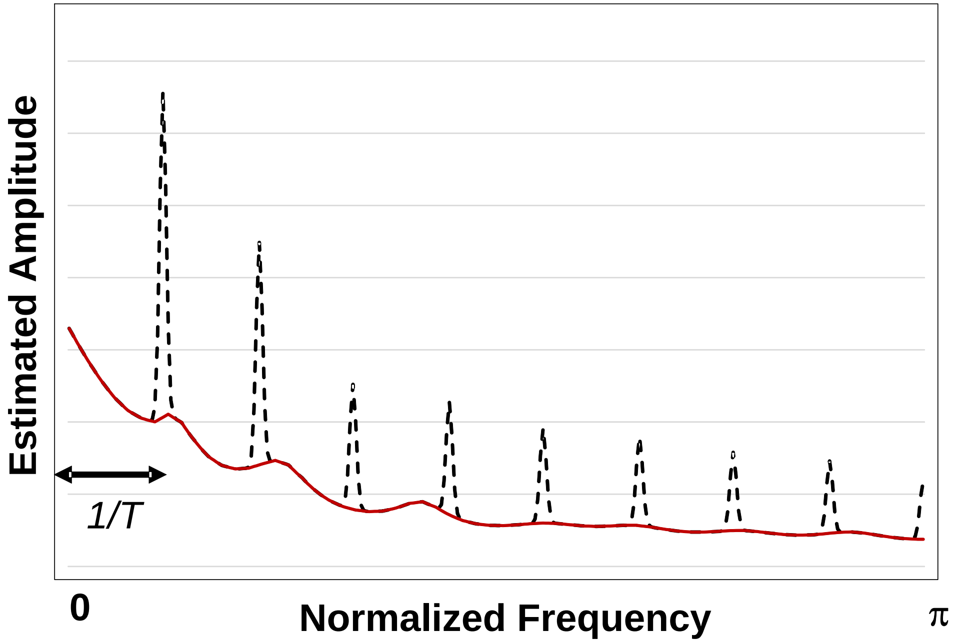
EVS supports several optimized coding techniques for covering different bit-rates and contents

- Low amplitudes: **Trellis Coded Quantization**
  - Very low amplitudes and LSBs of SQ
- Harmonic components: **Harmonic Vector Quantization**
  - Peaks position and magnitudes of the harmonic tones are coded separated
- Noisy components: **Pyramidal Vector Quantization**
  - Optimal for Laplacian distributed source
- Generic coding scheme: **Scalar quantization + arithmetic coding**
  - Based on probability model estimation
  - Adaptive SQ deadzone based on estimate of tonality
  - Add a harmonic model for enhancing probability models

Optimized spectral coding

Illustration of the redundancy exploitation by the entropy coding for harmonic signals

- Probability model for the code is first estimated from the spectral envelope (red line).
- For tonal items, an harmonic model can be added to refine the estimate (dashed line).



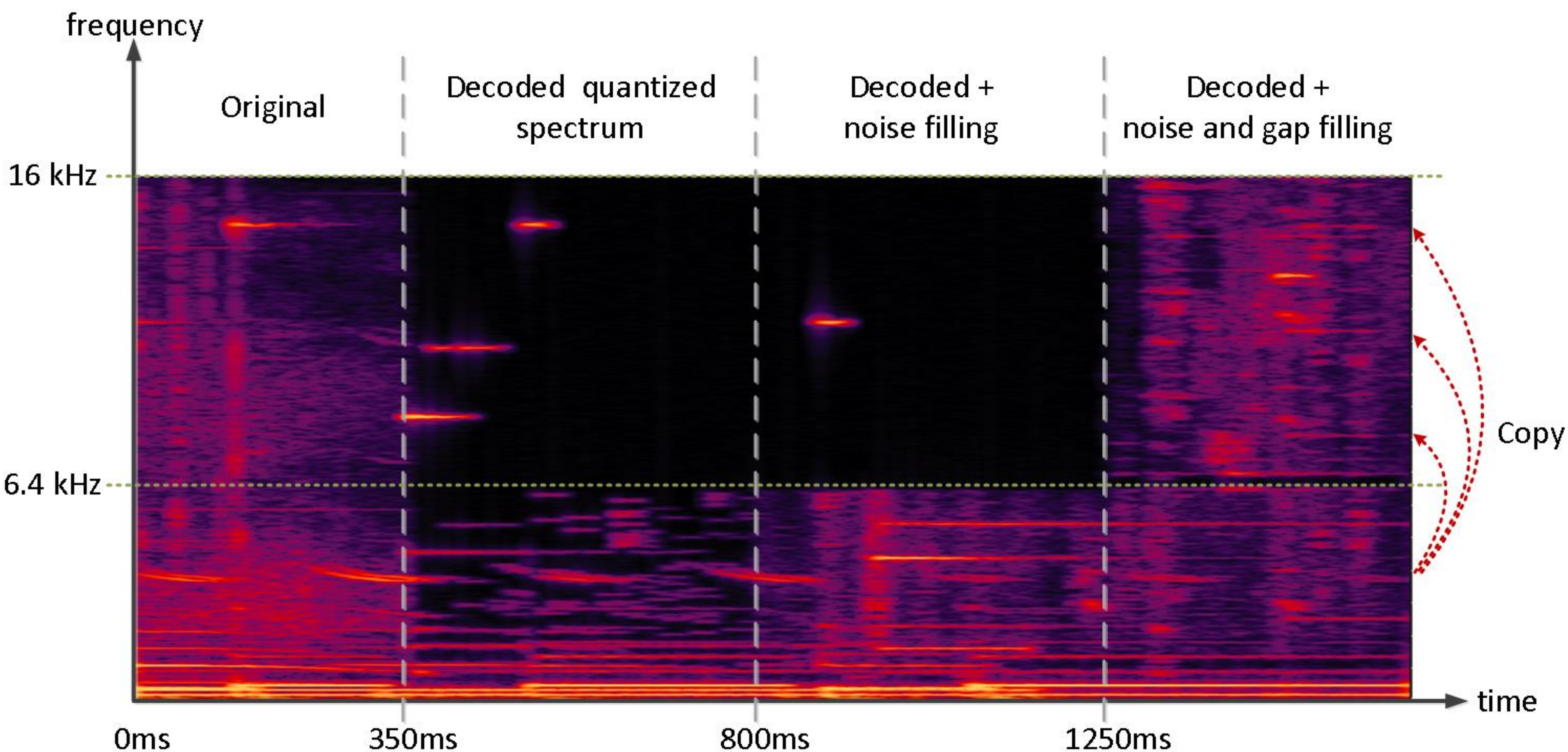
Noise and Gap filling

- Noise filling
  - Treat zeroed spectral lines by injecting random noise
  - Inserted noise is attenuated close to non-zero quantized lines
  - Avoids degradation of tonal components

- Gap Filling
  - Parametrizes least relevant high-frequency bands or zeroed frequency bands
  - Model using random noise and spectrum similarities
  - Shaped by the coded spectral and time envelope

Noise and Gap filling

Illustration of the effect of noise and gap filling



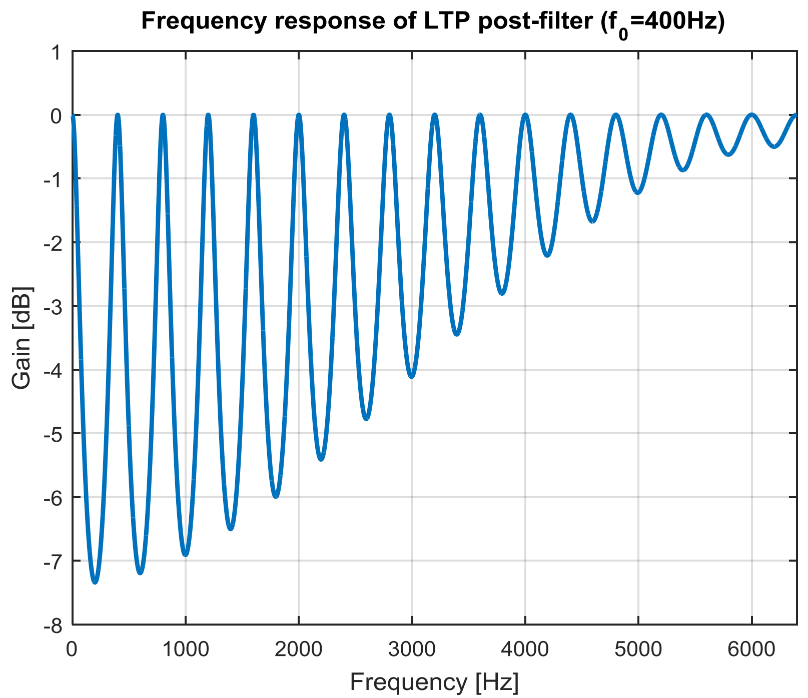
Concealment

Different concealment strategies are employed depending of the signal nature.

- **In Frequency Domain:**
  - Non periodic noise-like components: sign scrambling of the past spectral coefficients.
  - Tonal components: Phase prediction of the past sinusoidal components.
- **In Time Domain:**
  - Speech and single instrument music: excite a long-term and short-term predictive filters.
  - Very stationary signals: frame repetition with phase matching.

Post-processing

- Pre-echo attenuation
  - Attenuates energy of the decoded signal before onsets
  - Reduces the typical artefact due to quantization noise time spreading after the inverse transform
- Long-term predictive (LTP) post-filtering
  - based on the LTP delay
  - Controlled by a coarsely quantized gain
  - Principle similar to Bass-Post-Filter for speech coders
  - Enhances perceptually the harmonicity



Conclusion

Conclusion

- EVS is currently the most efficient and versatile codec for high quality communication in any type of network, including the VoIP and mobile networks
- Excellent performance in terms of compression and speech/audio quality
- Various new features, improvements and innovative approaches
    - switched speech/audio coding at low delay
    - wide range of operation points, stretching from highest compression to transparent coding
    - audio bandwidth up to 20 kHz
    - advanced compression efficiency
    - high quality for clean/noisy speech, mixed and music
    - high robustness against packet loss
    - AMR-WB IO mode for compatibility with existing systems

Slides will be made available at

http://www.aes.org/technical/cas/

ERICSSON Fraunhofer IIS HUAWEI VoiceAge®
www.voiceage.com