Next Generation Wifi VoIP Phones:

How to reduce the packet rate of speech codecs?

Dr. Christian Hoene
Universität Tübingen

ITG Fachtagung VoWLAN
Würzburg, May 15th, 2006

Content: First Part

- State of the Art
- VoWLAN Capacity
- Packet Dropping Strategy
- Limits of VoIP packet classification
- Real-time classification of µ-law frames
- Summary

Where are we now?

- Phy: IEEE 802.11b/g
- MAC: WMM or WME for QoS
- Security WEP-WPA2
- Network: IPv4
- Transport and Signaling: RTP, SIPv2
- Codecs: G.711, 723, 726
and G.729a/b
- Acoustics: Comfort noise generation (CNG), voice activity detection (VAD), adaptive jitter buffer and echo cancellation
- Battery life: With 800mA battery-Four hours talk time-Sixty hours stand-by time

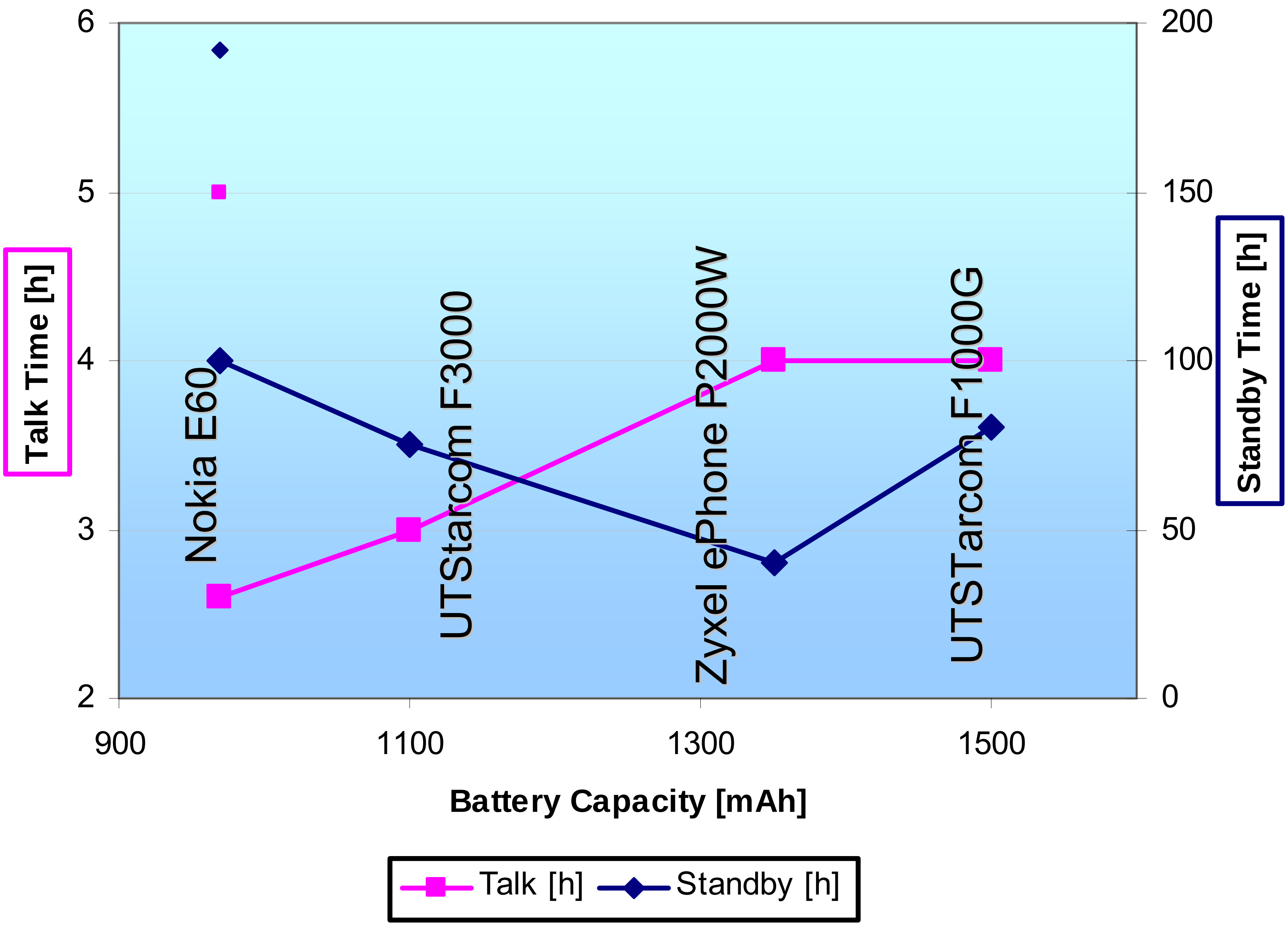
Comparing VoWLAN with GSM and DECT Phones

- GSM technology is used for cellular networks
  - e.g. a recent Nokia GSM/UMTS phone
  - Weight 117g
  - Battery: BL-5C Li-Ion (970mAh)
  - Standby time: 144 - 288 h
  - Standby time: 3,3 - 6,4 h
  - Transmission range: up to 35 km
  - Price: 340 €

- DECT technology is used for cordless phon
  - e.g. Siemens Gigaset S44
  - Weight 120g
  - Battery: NiMH AAA (750 mAh)
  - Standby time: up to 150h
  - Talk time: up to 10h
  - Range: up to 300 m
  - Transmission Range: 50 m (indoor) and 300 m (outdoor)
  - Price: 70 €

- Is VoWLAN a better technology than GSM or DECT?

Operating Times of VoWLAN Phones



VoWLAN Capacity depends on Frame Rates

- IEEE 802.11b has a high *packet switching overhead.*

- *VoIP transmission includes*
  - *RTP, UDP, IP headers*
  - *IEEE 802.11 MAC headers*
  - *IEEE 802.11 Physical preamble*
  - *Collisions and Contentions*

- *Capacity of VoWLAN depends on the number of packets per seconds*
  - *As various performance simulations and experiments have shown.*

- *Question: How to reduce the number of packets per second?*

Background: Adaptive Coding Rate in UMTS

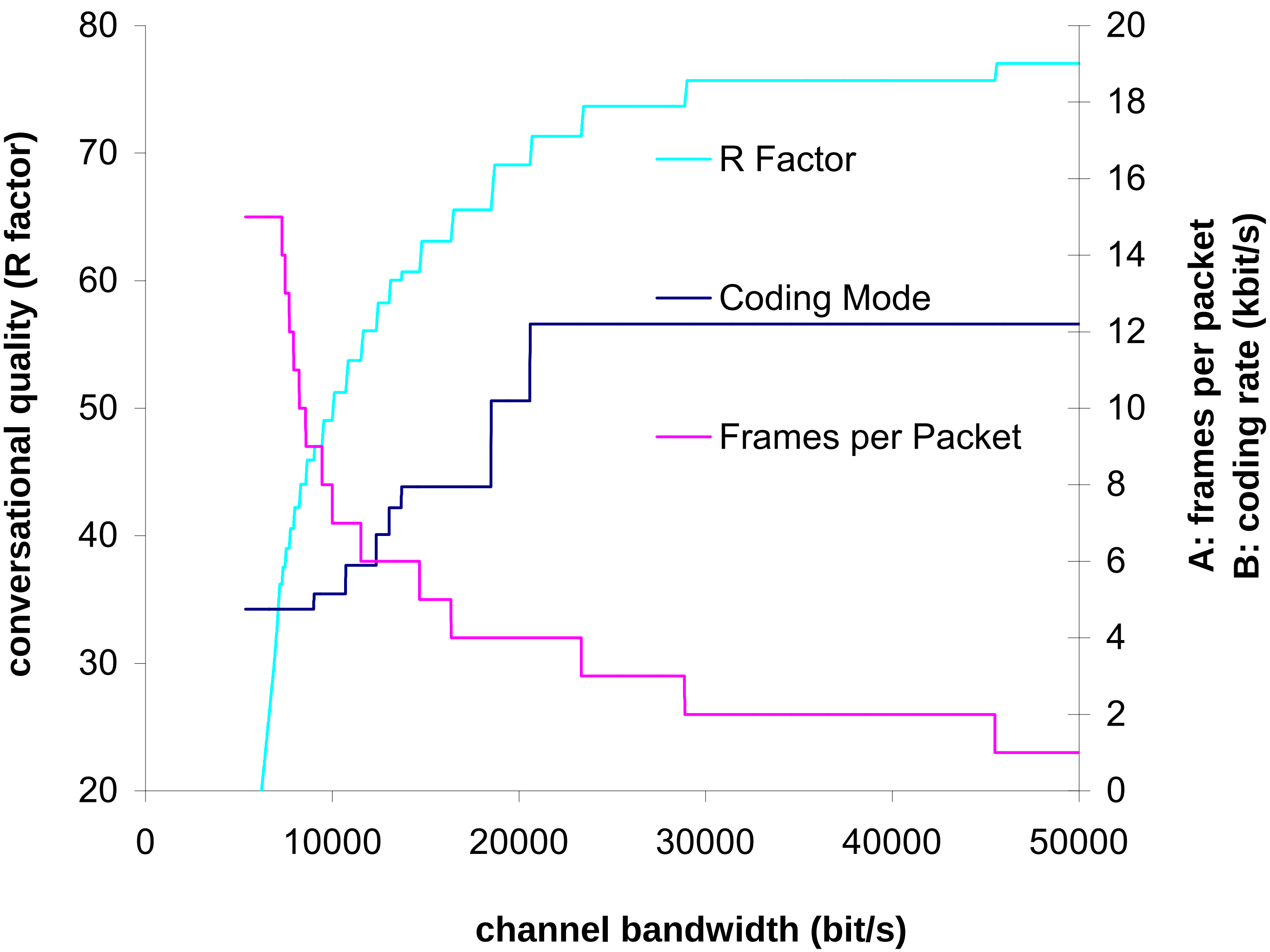
- Adaptive Multi-Rate Coding (AMR) supports
  - Eight different coding modes
  - Ranging from 4.75 to 12.2 kbps.
- Depending on the quality of the channel
  - The ideal coding mode is selected

- If the capacity is plenty
  - AMR-WB can be used (twice the frequency bandwidth and coding rates)
- If the capacity is low
  - Half-rate modes are used.

How to Measure VoIP Service Quality?

- Measuring Networking QoS
  - E.g. loss rate and mean delay
  - Easy but inaccurate
- In the end, the service quality is important
  - Human based listening tests are extensive
- ITU P.862 (PESQ algorithm) measures speech quality
  - Compares original sample with the transmitted version
  - Calculates Mean Option Score (MOS) (1=bad, 5=excellent)
- ITU G.107 (E-Model) predicts quality of tel.-system
  - Considers echo, loudness, coding, packet loss rate, delay, …
  - Result: R Factor (0=bad, 70=toll quality, 100=excellent)

Limited Bandwidth on an Ethernet-Link [1]

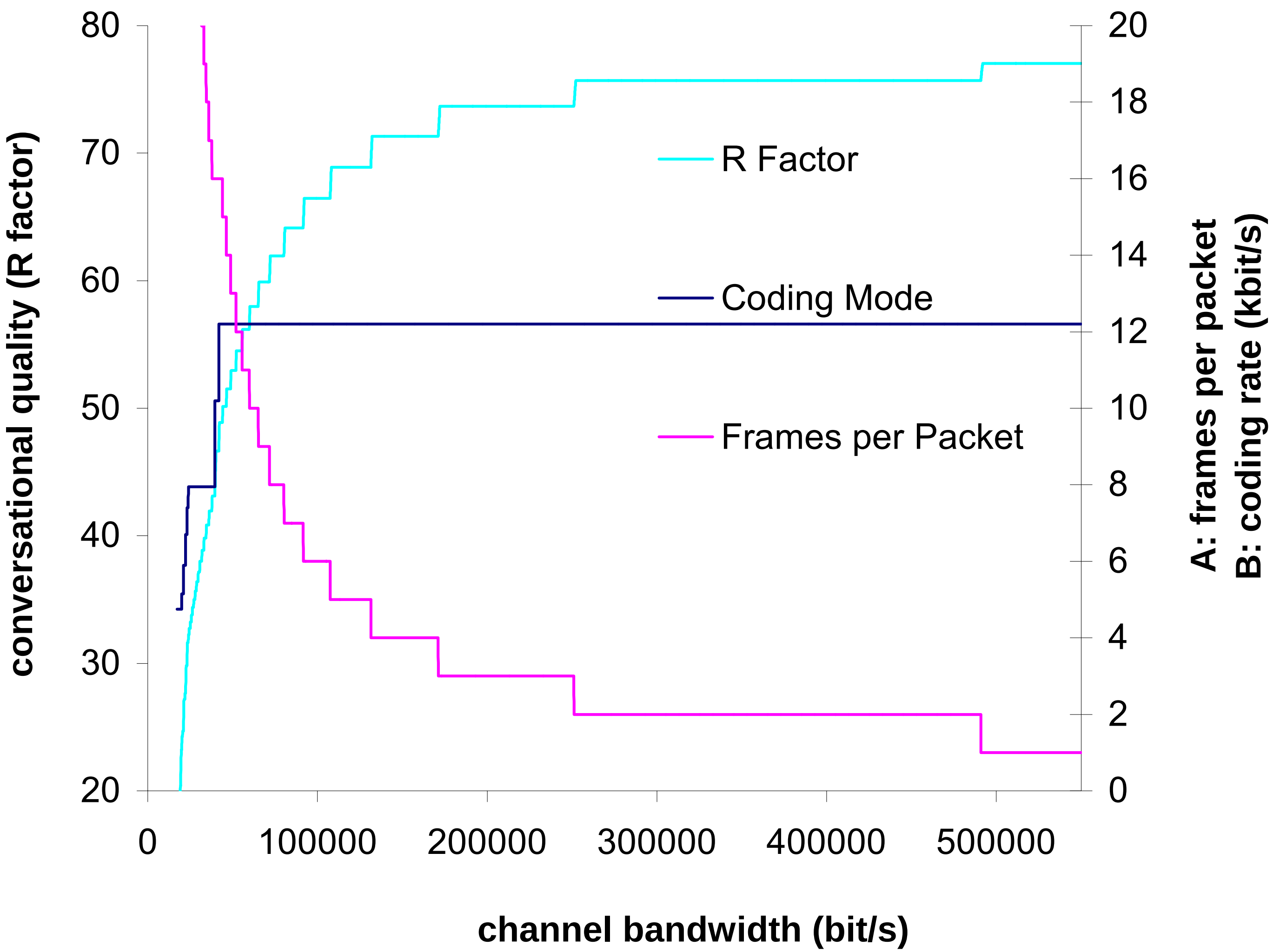


Overhead of sending a packet on IEEE 802.11b

Table 9.3.: Overhead in microsecond and bytes per UDP/IP datagram for the IEEE 802.11 platform as given in [8].

Mode [Mbps]	Packet headers, 82b [μs]	ACK overhead, 8b [μs]	Physical over-head [μs]	Overall [μs]	Overall [b]
11	59.6	56	754	869.6	1196
5.5	119.3	56	754	737.3	639
2	328	56	754	946.0	284
1	656	112	754	1522.0	190
Consists of	RTP[12b], UDP/IP[28b], SNAP[8b], MPDU[30b], FCS[4b]=82b	ACK[8b]	2*PLCP[192μs], DIFS[50μs], SIFS[10μs], Backoff[310μs]	Packet headers, ACK, Physical overhead	Packet headers, ACK, Physical overhead

Limited Bandwidth on IEEE 802.11b



Theoretical Results

- If the wireless link is congested,
  a Wi-Fi VoIP phone shall adapt
  - Packet rate
  - Not coding rate!
  - (only if already a low-rate coding is already applied.)

- Why?
  - Because every VoIP packet has a large overhead
    - due to packet headers (e.g. IEEE, LLC, IP, UDP, RTP)
    - due to MAC overhead (e.g. contention and collisions)

Problem: How to reduce the packet rate of speech coding?

1. Packetization:
   - Change the number of speech frames per packet.
   - Is supported by current standards (RTP)
   - Drawback: Increase the packetization delay.
2. Change the codec:
   - Currently, each codecs sends out frame at fixed rates (e.g. 10, 20, 30ms)
   - Why? Because of circuit switched transmission channel (ISDN, GSM, UMTS, …)
   - Current speech codecs have been design with CS transmission in mind.
   - No “real” Internet codec available
     - Even iLBC and other GlobalIPSound’s codec have fixed frame rates.

   - We need to **optimize the frame rate** in addition to
     - Coding rate, complexity, delay, and speech quality!

3. Suppress not relevant frames and do not transmit them..

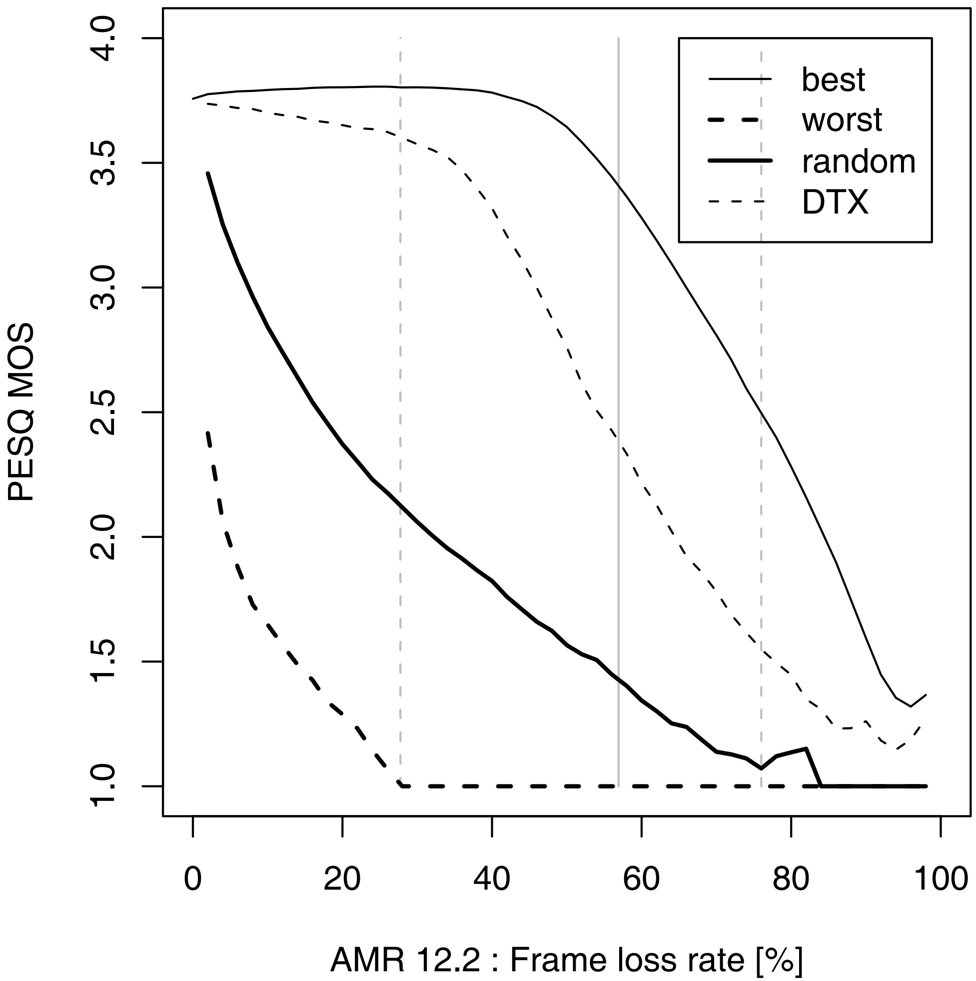
Dropping frames.

- Discontinuous Transmission (DTX)
  - Speech frames during silence are less important
  - Lower frame rate during silence

- But even frames during voice active differ.
  - Some frame are more important than others

- We have had developed an approach on how to measure the
  - Frame importances alias
  - The impact of a frame’s loss on speech quality

Speech Frames – Dropping Strategy



Drop packets in cases of

- ❑Congestion
- ❑Wireless fading
- ❑Saving transmission energy

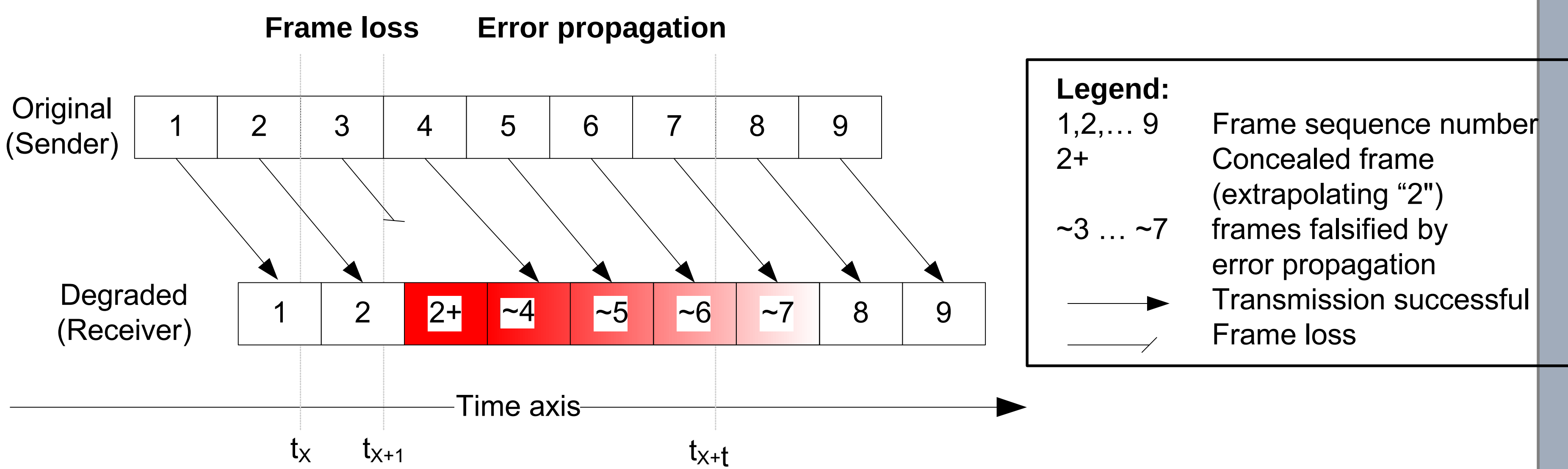
- ❑**Best**: dropping the unimportant frames first
- ❑**Worst**: dropping the important frames first
- ❑**Random** frames losses
- ❑**DTX**: drop first silences frames, then active frames (randomly)

Problem Statement

- Frame importance be can measured offline (previously presented approach).
- Offline not useful for interactive telephony!

- Can we predict the importance at real-time?
- How well perform other frame classification algorithms (benchmarking related work)?
- Can be we provide a better solution?

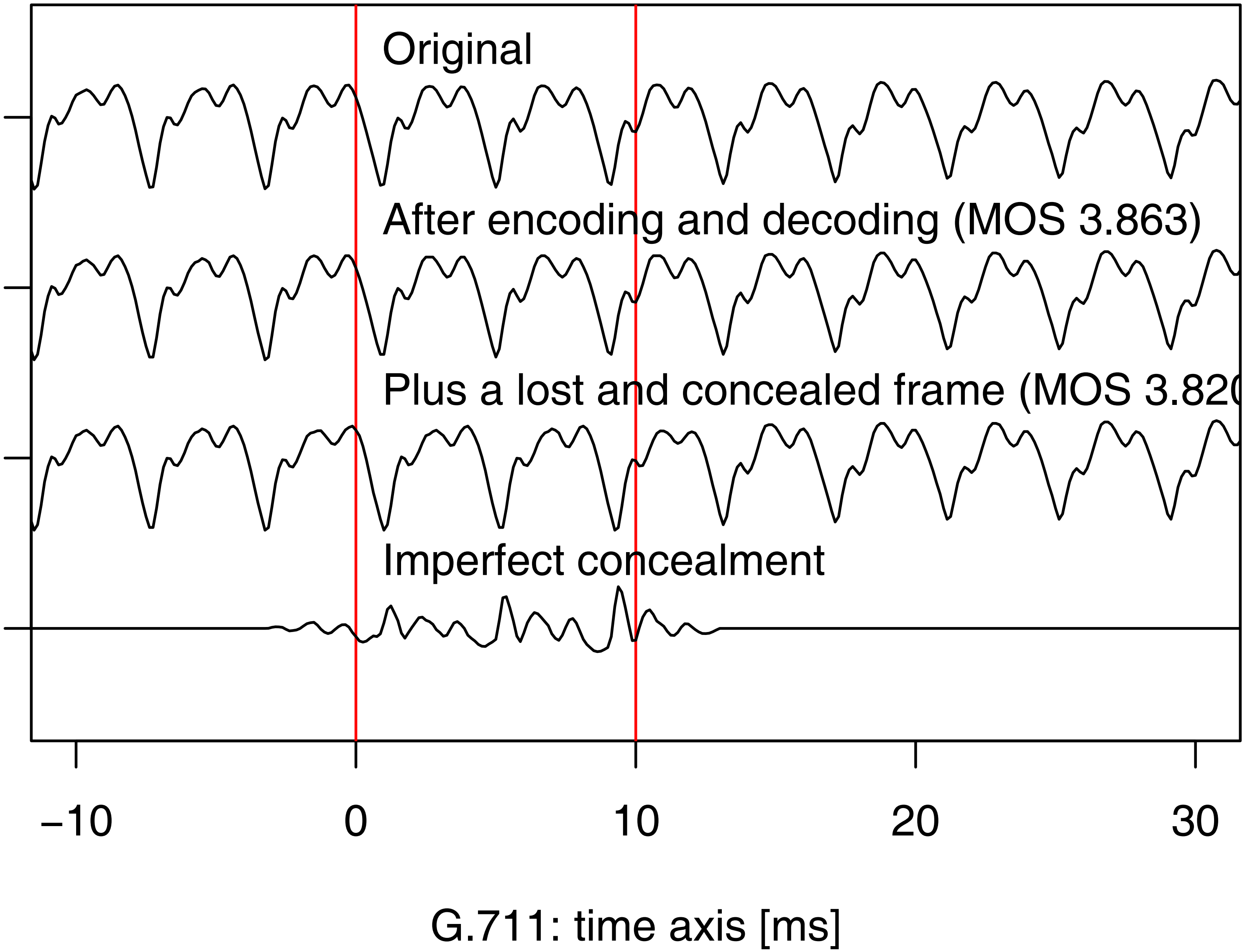
Understanding the Importance



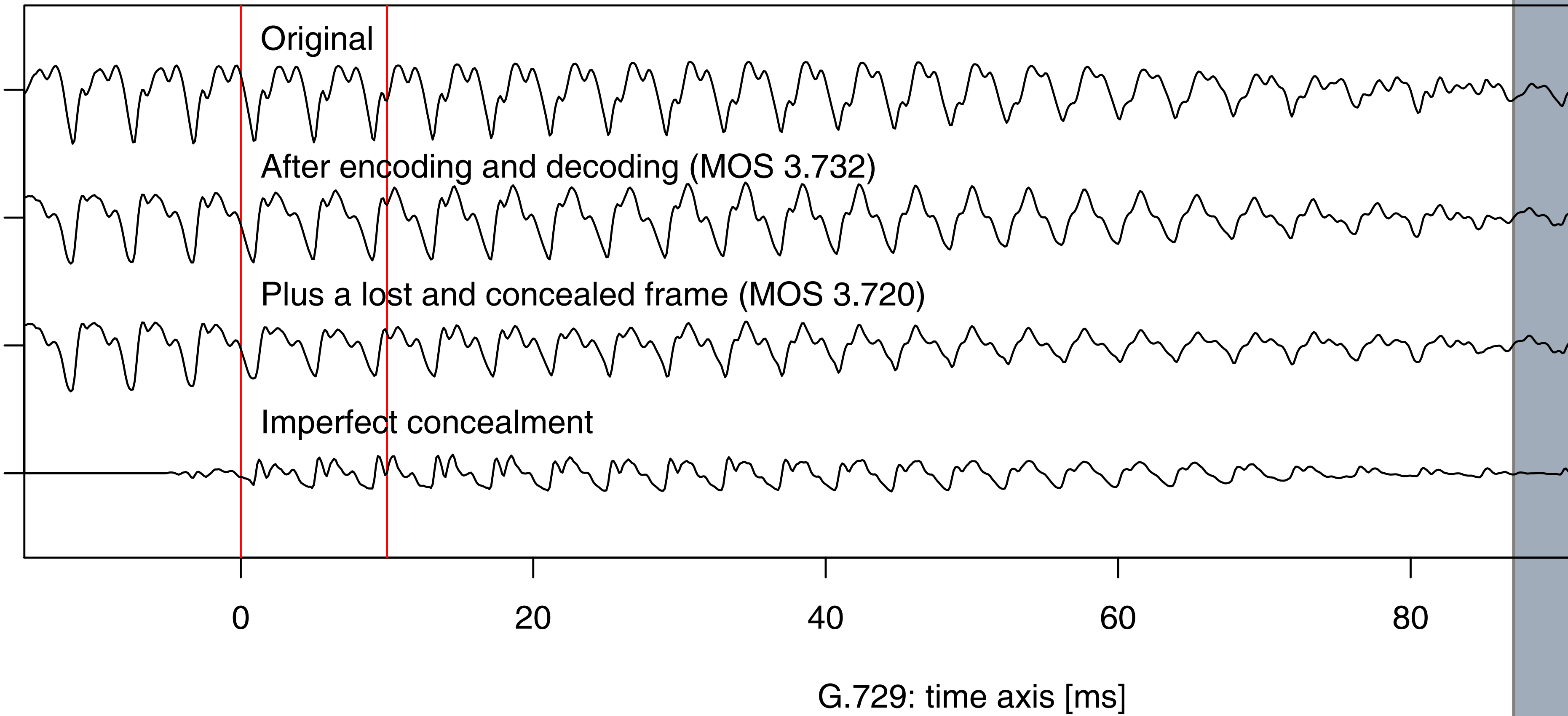
Frame loss distortion is due to two effects

- Imperfect frame loss concealment (2+ ≠ 3)
- Error Propagation (4…7 ≠ ~4…~7)

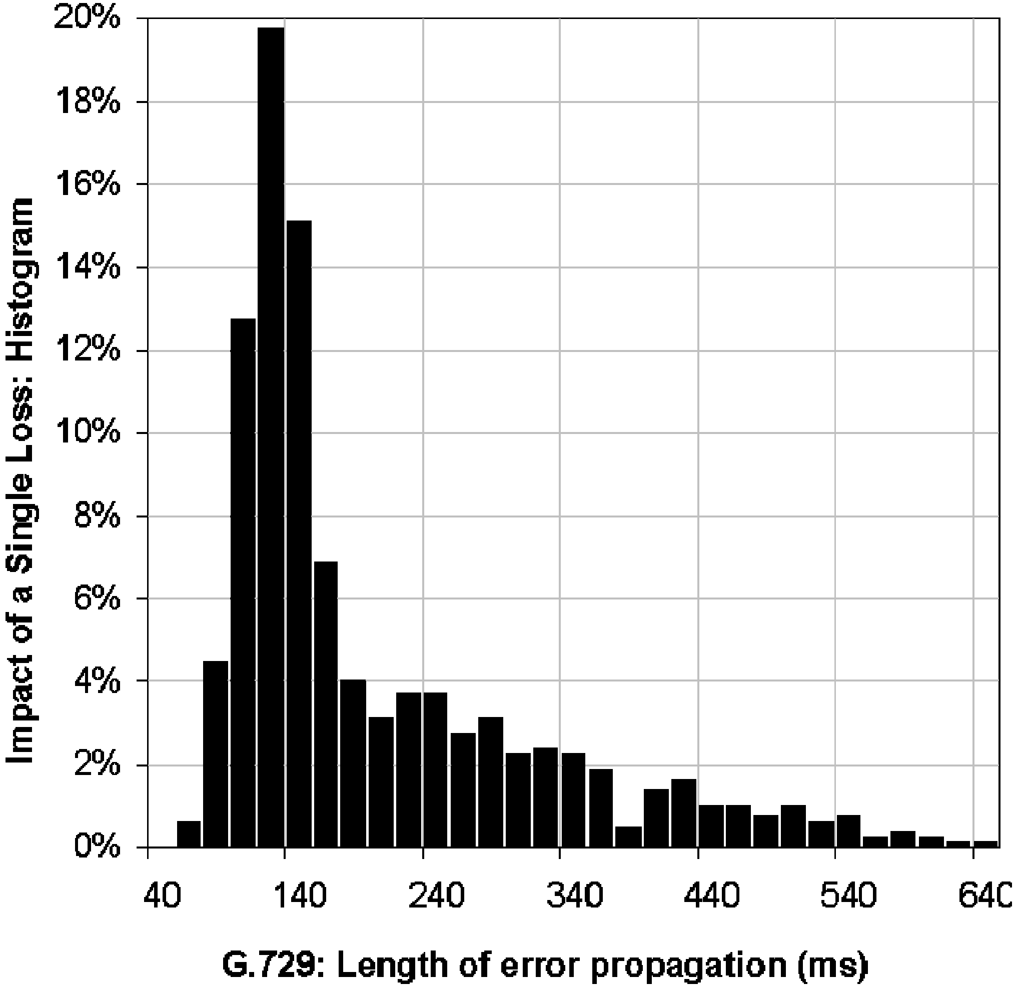
Example: One Loss with G.711 µLaw



Example: One Loss with G.729 Coding



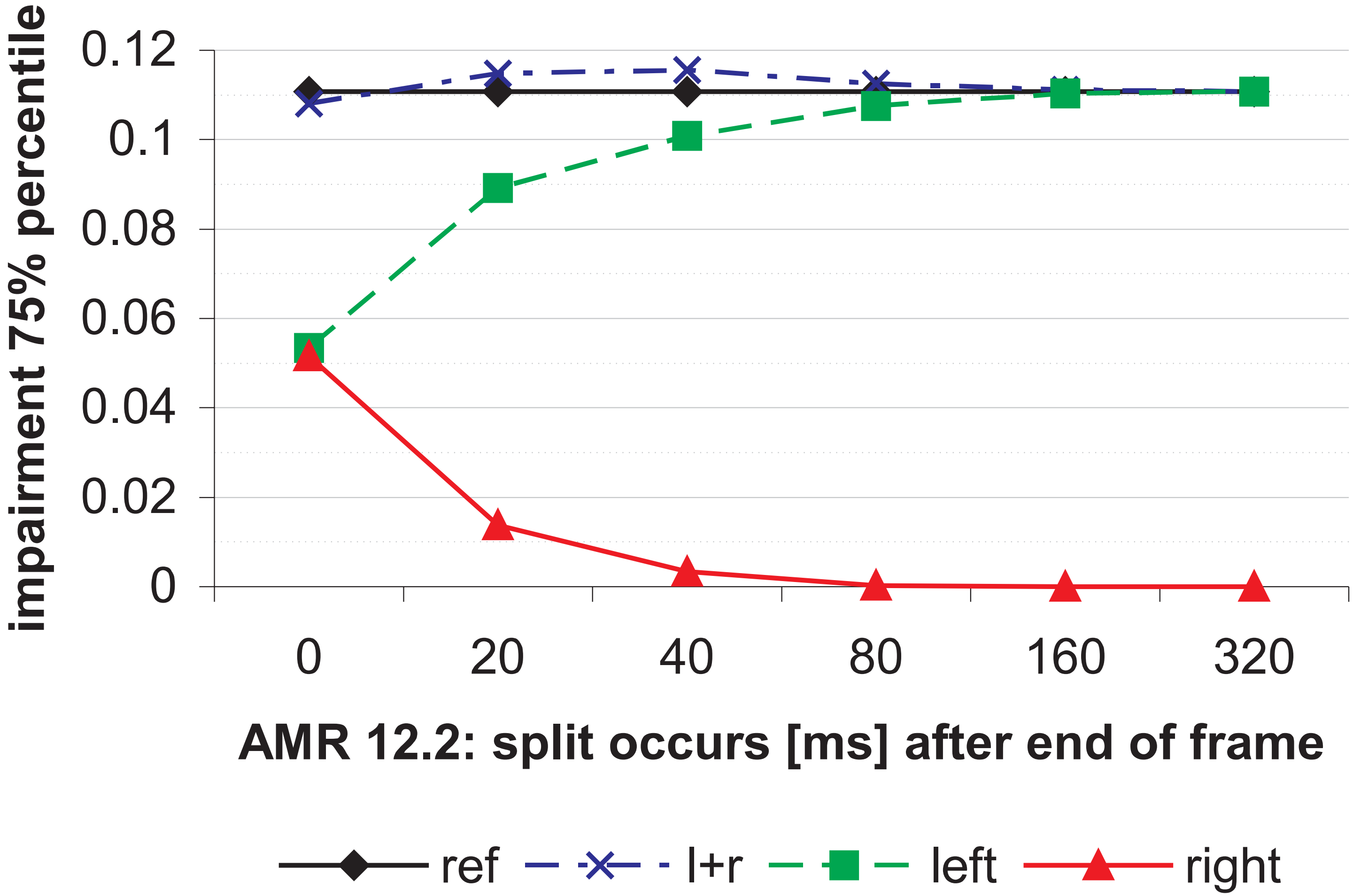
Length of Error Propagation After a Frame Loss



- Comparing internal decoder states of no-loss with the loss case
- and measuring the length of the mismatch.

- (ignoring decoders post filter as it never comes back to normal.)

Temporal impact of error propagation



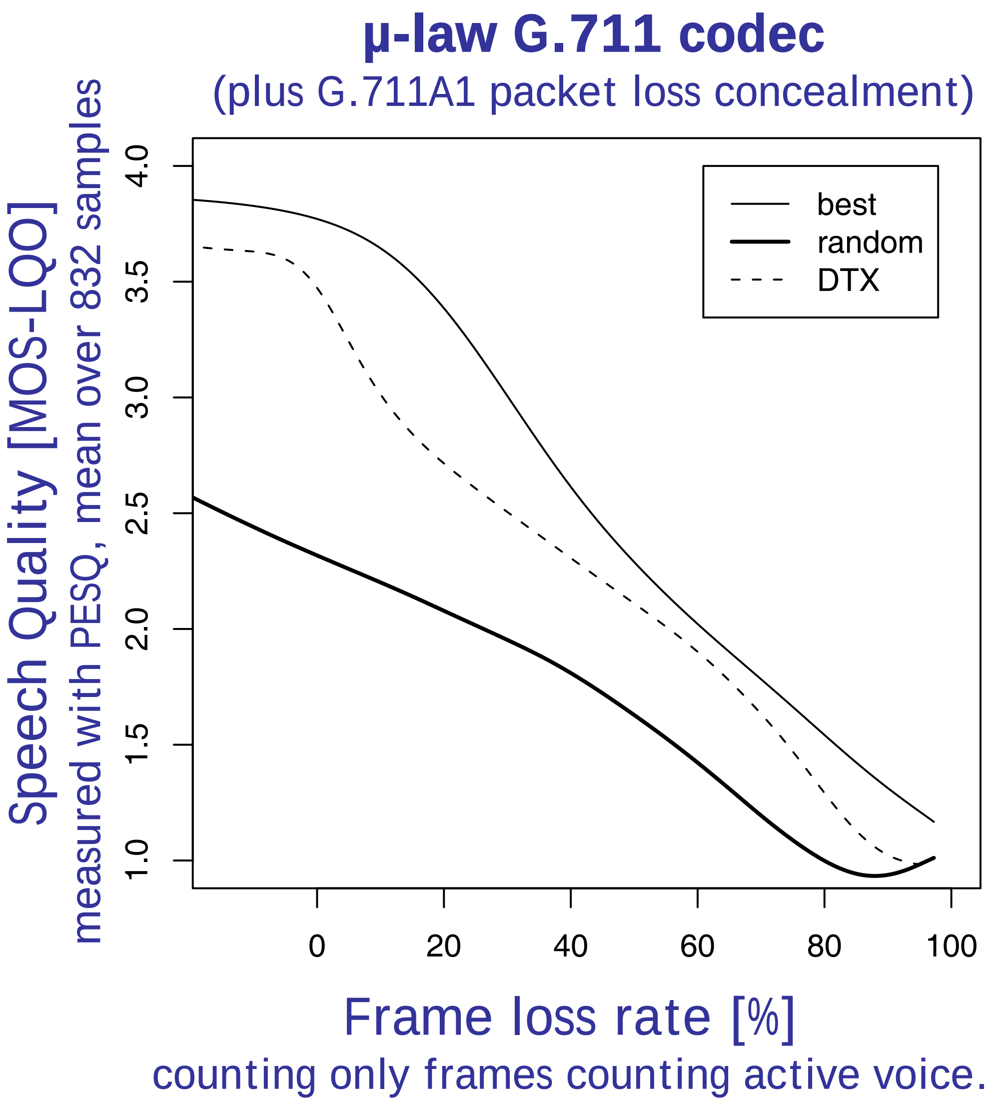
- Concealment (left) vs. Error-Prop. (right)

Analysis of Prediction of Packet Importance

- Silence detection of G.729 and AMR encoder
- Voicing decision of G.729 and AMR decoder
- De Martin‘s Packet Marking: re-implemented
- Sanneck SPB DiffMark: Software available

- Analysis:
  Correlate importance of a packet as measured with PESQ with the packet loss prediction algorithms

The Importance of Speech Frames Differs



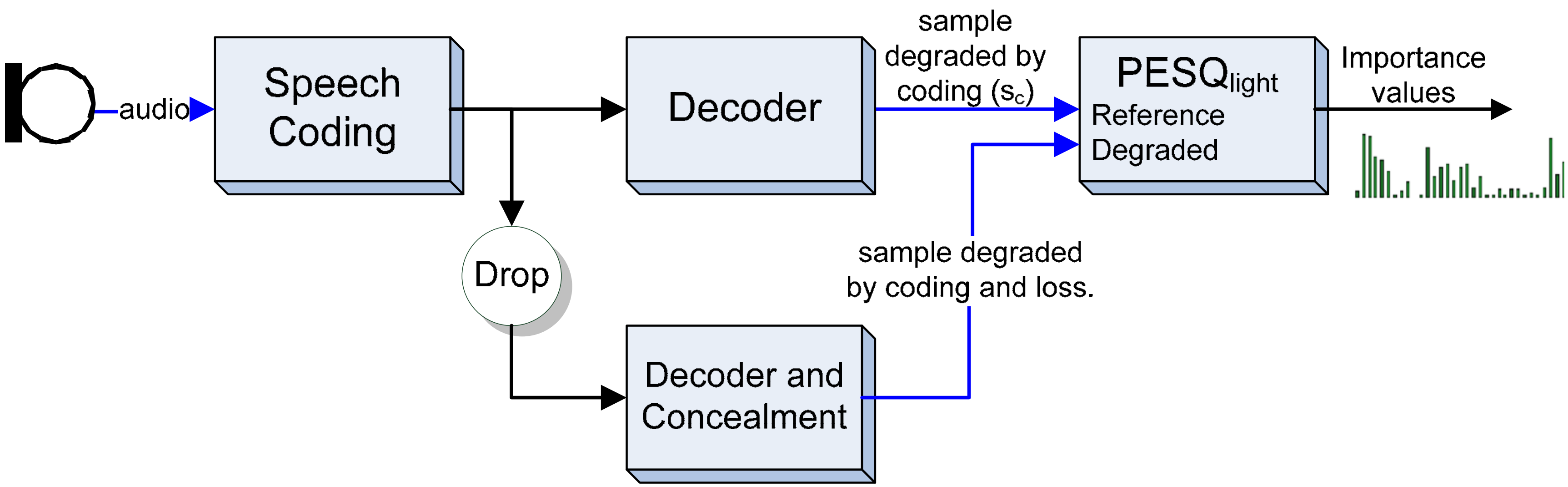
- Some frames are important
- Many other frames during voice activity can be dropped without any large impact on speech quality.

Impact of Frame Dropping

- **Best**: dropping the unimportant frames first
- **Random** frames losses
- **DTX**: drop first silent frames, then active frames (randomly)

- Previous work determined the **importance of frames** in an **offline** approach.

Real-time Classification of μ-law Frames



Schematic of a sender-side calculation of the importance values.

- Simulation of receiver at the sender (analysis-by-synthesis approach)
- Using a perceptual quality model (PESQlight)
  - based on PESQ
  - but reduced computational complexity by a factor of 3.5.
- Using only short pre-leading and following periods of speech.
- Tradeoff between complexity and quality of prediction. **Calculation in real-time.**
- Performance: Correlation of up to **R=0.63** with the reference importance values.

PESQlight (thanks to Till Wimmer)

- ❑Work in real-time on a notebook.

Similar to offline approach but

- ❑Uses one MOS calculation instead of two.
- ❑Uses 250ms in front and 0 or 20ms after loss.
- ❑Uses complexity reduced version of PESQ (work of my student Till Wimmer).
- ❑Further optimization possible…

Table 7.4.: Functionality removed the original from PESQ algorithm.

Function	Description	Function	File
Time alignment	The time alignment can be removed, because between original and disturbed signal no variable delays are introduced.	input_filter, calc_VAD, crude_align, utterance_locate	pesqmain.
Voice Activity Detection	The Voice Activity Detection (VAD) occurs naturally already before calculation of the importance values. For example, a VAD is included in the encoder, the adaptive gain control or the echo compensation.		
Power reference	The values are not required for the real PESQ functionality.	pow_of	pesqmod.
Utterances	The subdivision into several utterances is not necessary, because only speech segments not larger than one second are considered.	short_term_fft	pesqmod.
Frequency responses compensation	No constant frequency distortion is to be expected because of the given codec.	total_audible, time_avg_audible_of, freq_resp_compensation	
Constant loudness	??	fix_power_level	pesqmain.
Skip silent sam-	??		pesqmod.

Final Contest

Algorithm	Correlation Coefficient (R)
G.729 Voicing on G.711	**0.184**
De Martin G.729 (on frames)	**0.195**
De Martin G.729 (only unvoiced frames)	**0.469**
SPB-Diffmark	**0.104**
Our algorithm (seg. length 0.25s, dropped frame is the next to last frame)	**0.600**
Our algorithm (seg. length 0.25s, dropped frame is last frame)	**0.318**

Next research steps in speech coding:

- Develop an online classification schemes for codecs like AMR, AMR-WB, AMR-WB+
- Develop codecs that scale with bit rate to perfect quality to utilize broad band transmissions
  - Push standardization?
- Develop codecs with source and flow adaptive bit- and frame rates.
- Develop transport protocols for next generation VoIP codings.

Benefits of low-frame rate codings:

- Increase capacity if bandwidth is limited
- Decrease mean energy consumption if powered by a battery

The End

- Thank you for your attention!