Error Propagation After Concealing a Lost Speech Frame

Christian Hoene (University of Tübingen)
Ian Marsh (KTH Stockholm, Sweden)
Günter Schäfer (University of Illmenau)
Adam Wolisz (Technical University of Berlin)

MULTICOMM Workshop
Istanbul, 11. June 2006

Introduction

Losing one Voice-Over-IP frame impairs

the perceptual quality

in a wide range, depending on

- the frame speech properties
- the encoder/decoder/concealment algorithms

- **decoders resynchronization time after loss (especially low-rate decoders might maintain a wrong state after loss lasting for the following frames.)**
- **the surrounding/following speech.**

Additive Metric to quantify the Relevance of Speech Frames

Definition:

The importance of frame losses is the difference between the speech quality due to coding loss and the quality due to coding loss and frame losses, times the length of the analyzed sample:

$$Imp(s,c,e) = (cl - c)\cdot t(s)$$
$$\text{with } cl = (4.5 - MOS(s,c,e))^2 \text{ and } c = (4.5 - MOS(s,c))^2$$

s: sample

t(s): samples length (s)

c: codec implementation

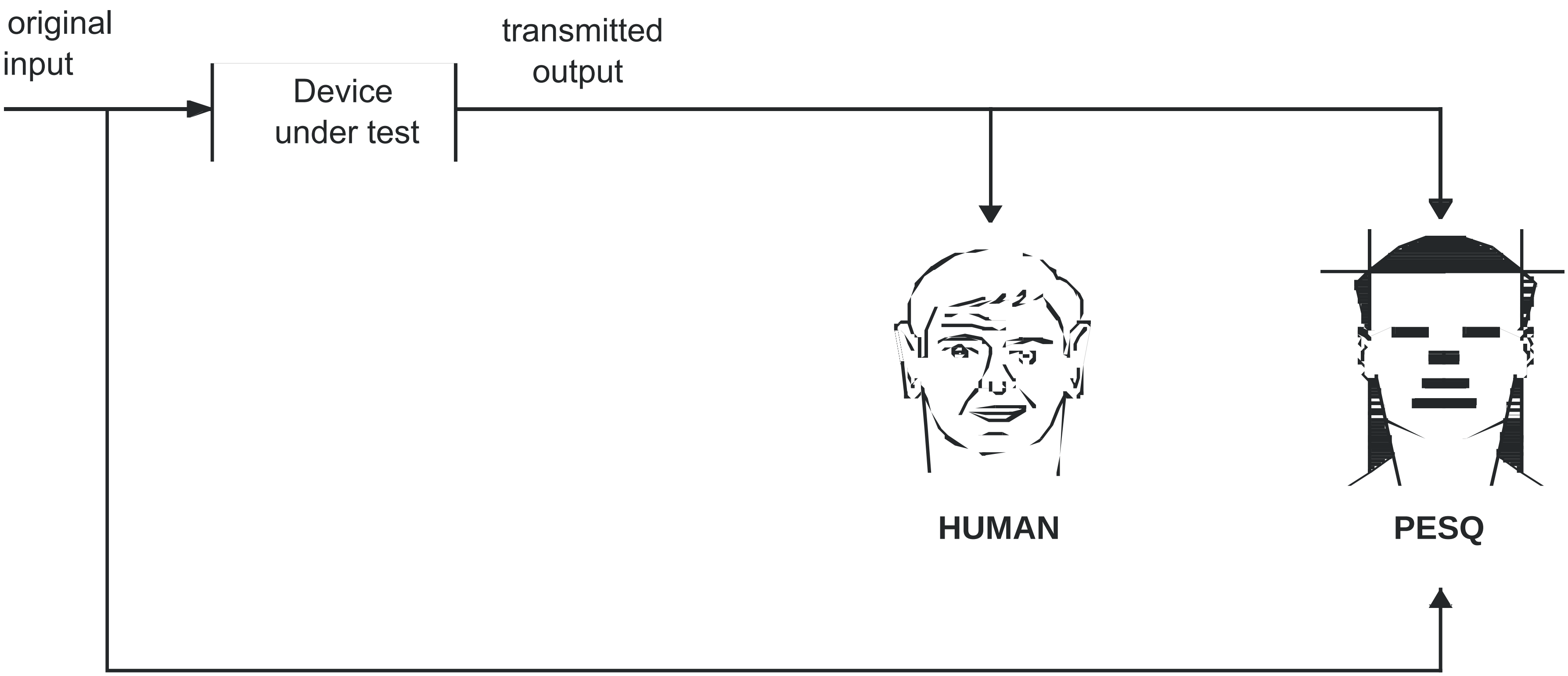
e: loss event, one or multiple correlated frame losses

- This equation remodels the behavior of ITU P.862 PESQ algorithm.
- We use a similar algorithm for aggregating frame losses as PESQ uses to aggregate the distortion of speech signals.
- This metric scales linear with the distortion (to some limits).
- We can **ADD** frame importance values (as least if they are distant).

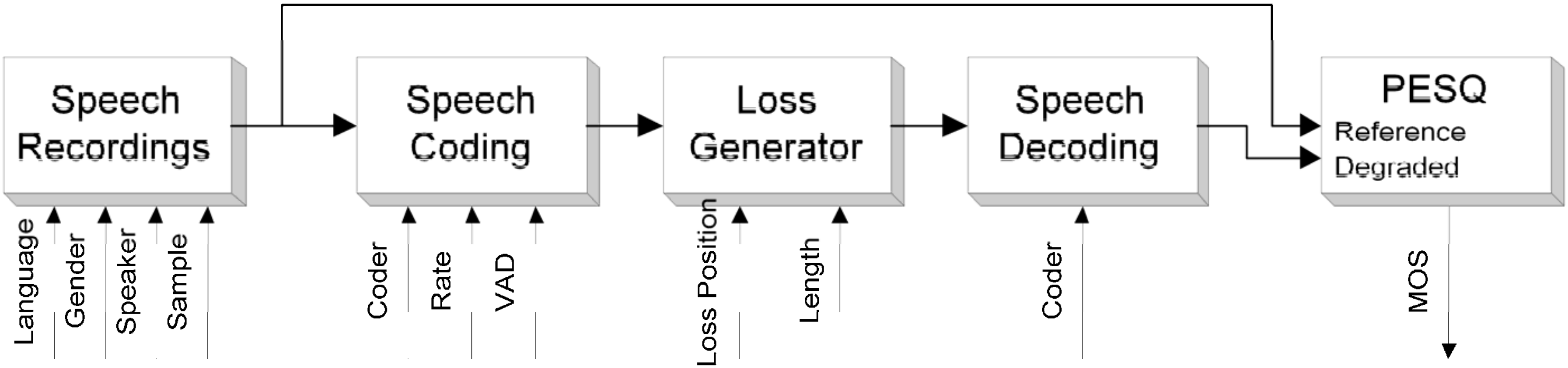
PESQ – Measuring Speech Quality.

How the measure the speech quality?

- Using formal listening-only tests (ITU P.800)
  - Human based listening tests are extensive
- ITU P.862 (PESQ algorithm) predicts human ratings
  - Compares original input with the transmitted version
  - calculates Mean Option Score (MOS) (1=bad, 5=excellent)



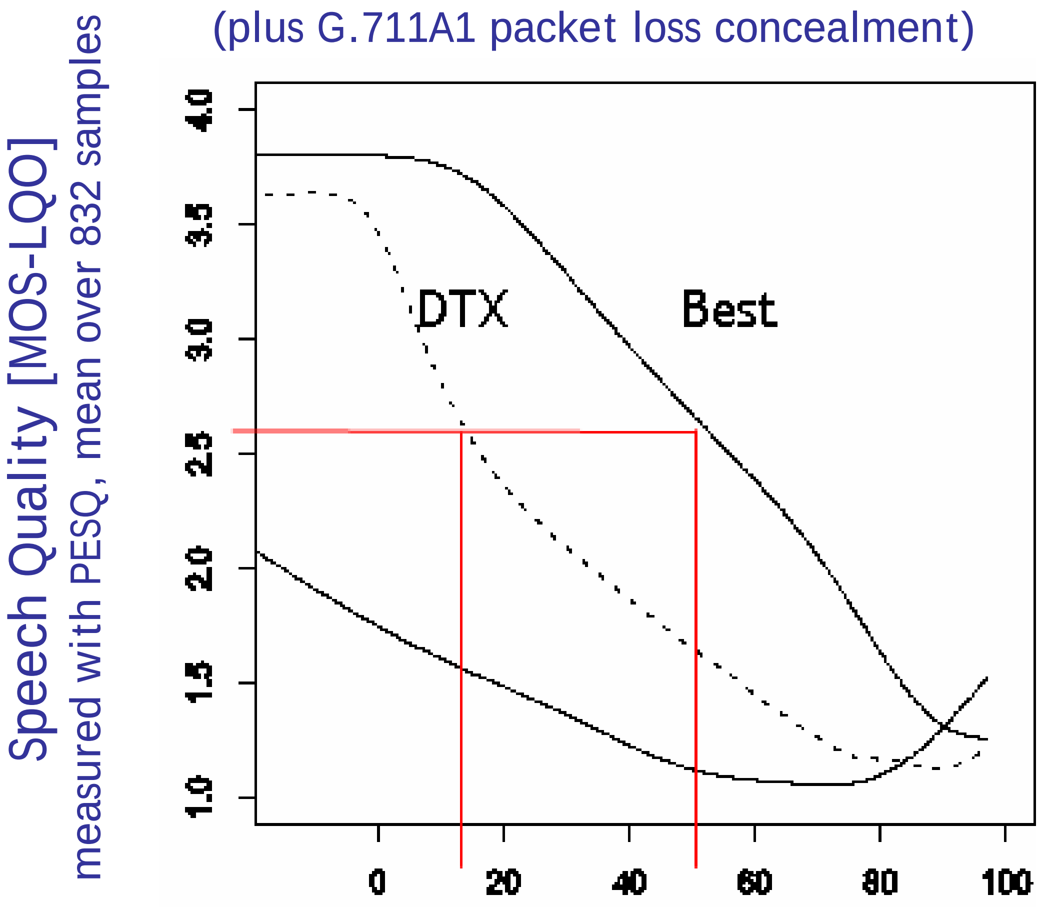
Collecting Statistics about Packet Losses



- Large sample database (ITU P suppl. 23)
  - 4 Languages x 4 speakers x 52 samples = 832
  - 8s each, two sentences
- Codec's:
  - ITU G.711 + Appendix II (64 kbit/s)
  - ITU G.729 (8 kbit/s)
  - 3GPP Adaptive Multi-Rate (4.75...12.2 kbit/s)
- Loss Generator
  - Different Positions (50) and Loss Lengths (1,2,3,4)
  - Totally: some million different single packet loss tests
- Use ITU P.832 PESQ to conduct tests.
  - PESQ calculates a Mean Opinion Source
  - Measurement procedure has been verified with formal listening-only tests (R=0.94)
- Just try it by yourself
  - www.tkn.tu-berlin.de/research/mongolia

The Importance of Speech Frames Differs

μ-law G.711 codec
(plus G.711A1 packet loss concealment)



Frame loss rate [%]
counting only frames counting active voice.

- Some frames are important

Drop packets in cases of
- Congestion
- Wireless fading
- Saving transmission energy

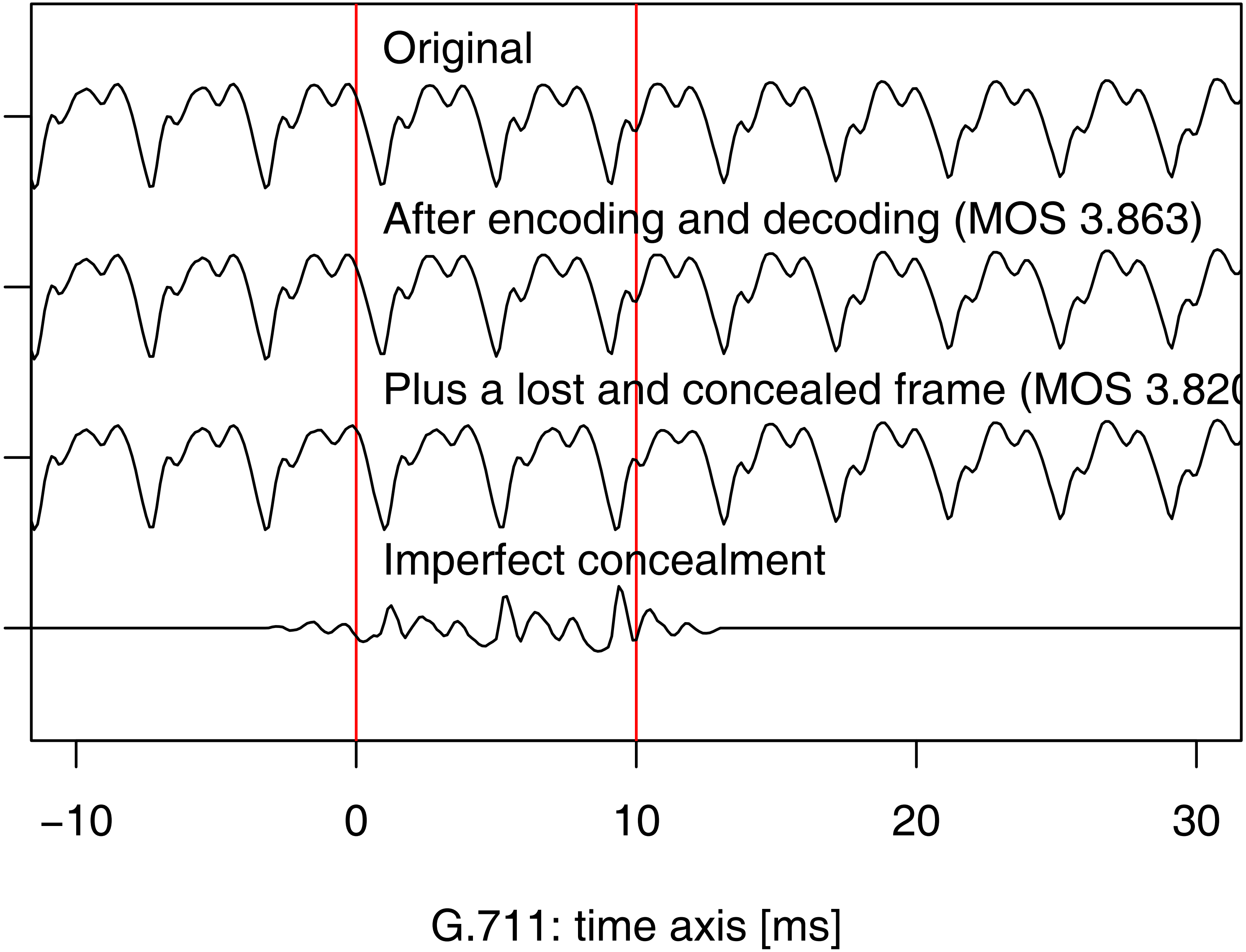
Impact of Frame Dropping
- **Best**: dropping the unimportant frames first
- **Random** frames losses
- **DTX**: drop first silent frames, then active frames (randomly)

Problem Statement of this Publication

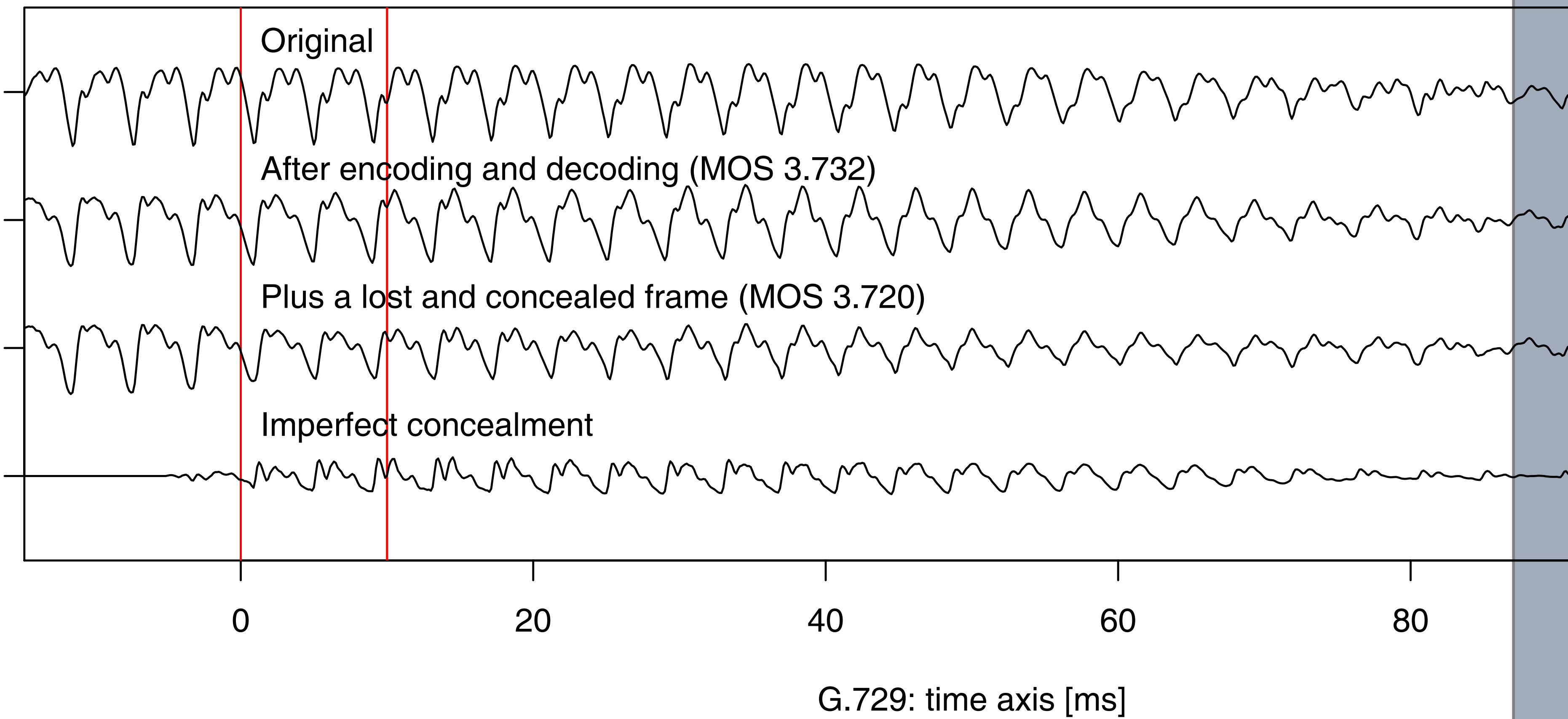
- Frame importance values be can measured offline (previously presented approach).
- Offline not useful for interactive telephony.
  - We need the importance at the time of transmission.

- Can we predict the importance at real-time?
  - to what extend?
  - Determine the limits of real-time packet classification!

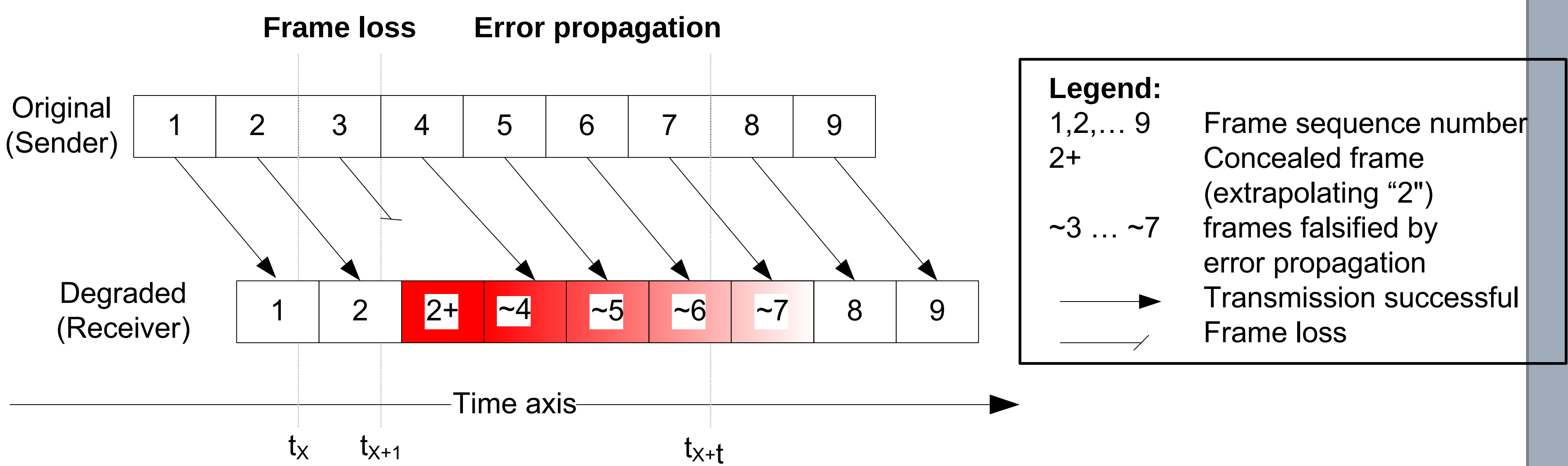
Example: One Loss with G.711 µLaw



Example: One Loss with G.729 Coding



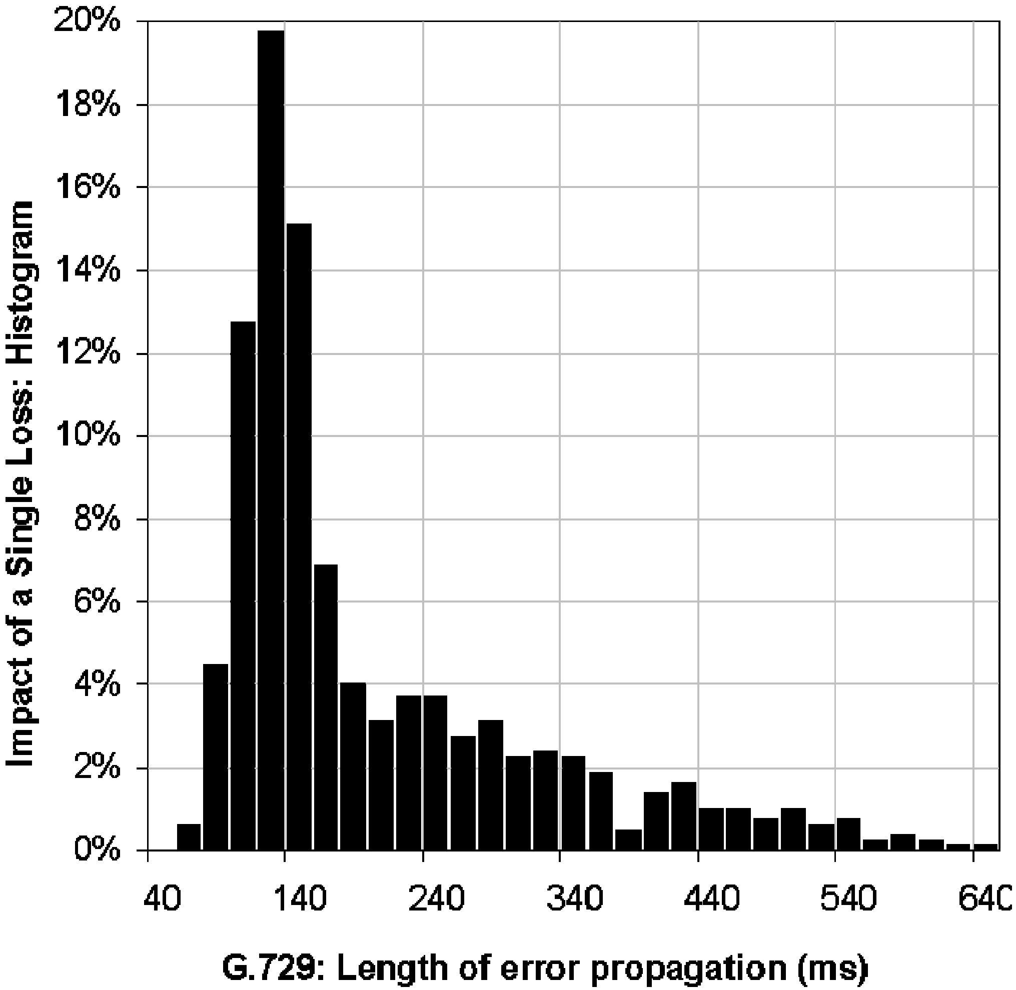
Understanding the Importance



Frame loss distortion is due to two effects

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

Length of Error Propagation After a Frame Loss



- Comparing
internal decoder states
of none-loss with the loss case

and measuring the length of the state mismatch called desynchronisation

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

How to quantify impact of error propagation?

- Non-trivial problem.
  - (It took me one year to solve it…)

- Measure it with PESQ to get perceptual relevant statement.

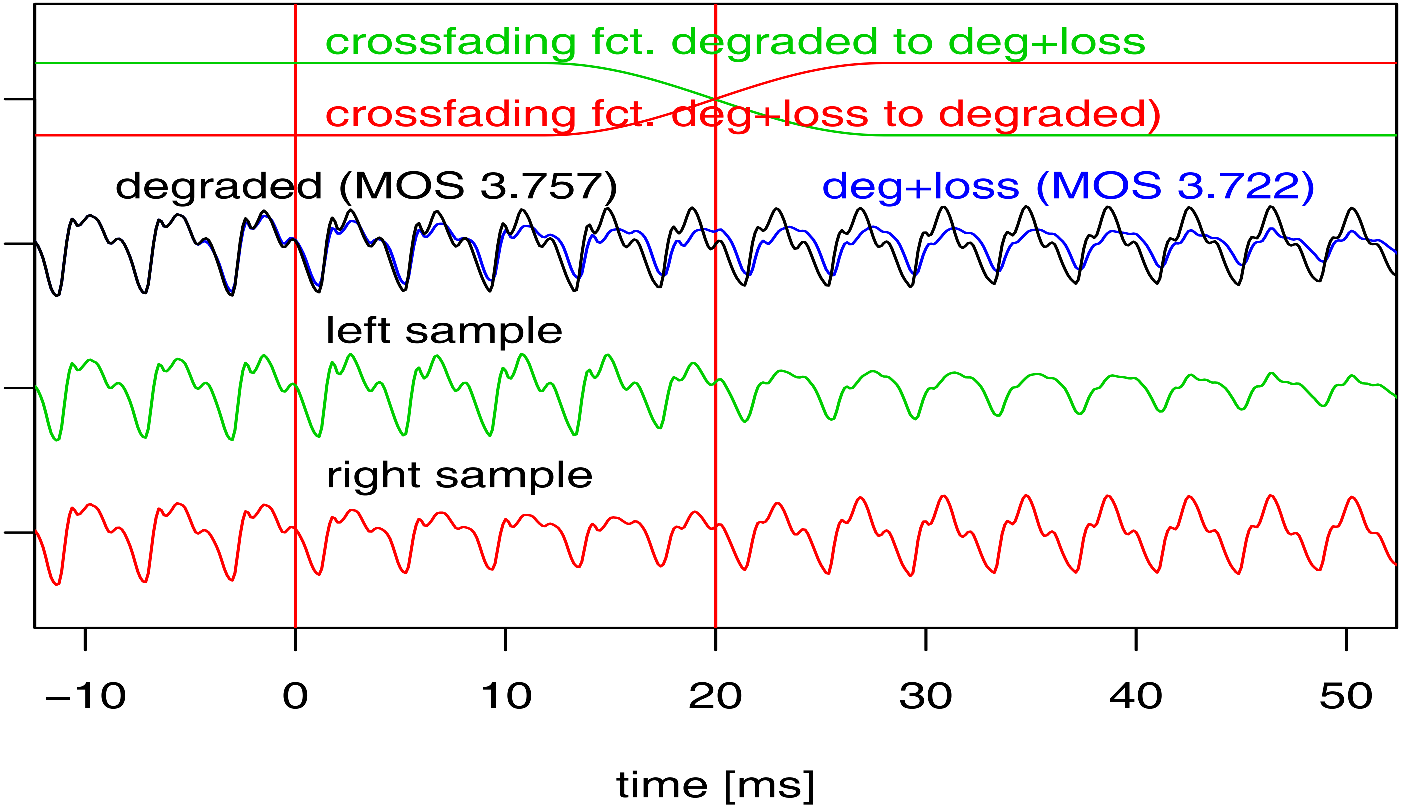
- Thus, do not split the samples before and after loss
  - This was my first try. It failed.

- Do not change the content of the sample,
  - because PESQ results depends highly on the content of the sample.

Slitting and Merging the Sample

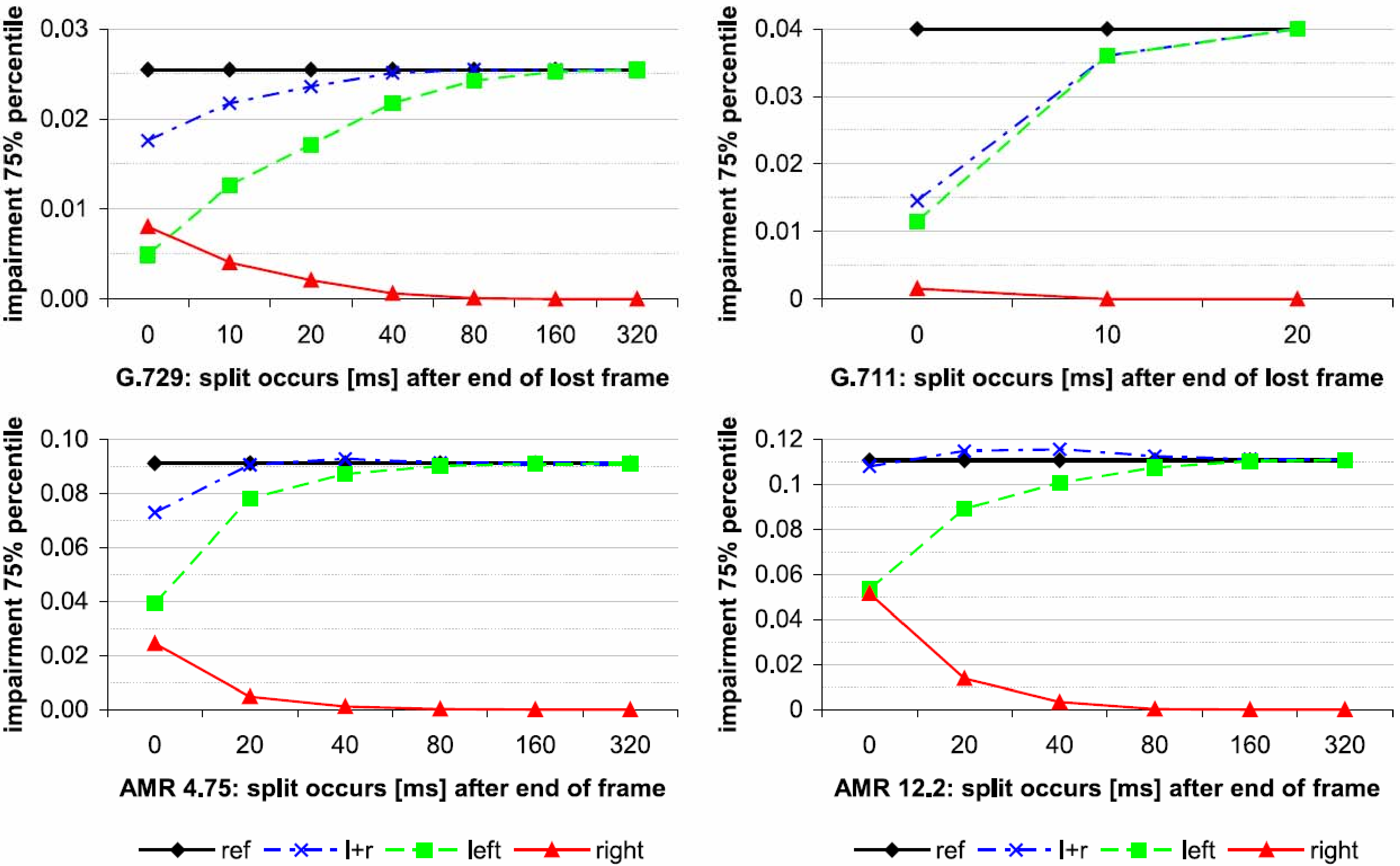
- Work THREE encoded sample
  1. Original
  2. Encoded
  3. Encoded and lost frames
- Slit sample 2 and 3 exactly after the frame loss
  - Sample 3-right part contains the effects of error propagation
  - Sample 3-left part contains the imperfect concealment
- Merge
  - sample 3-right with 2-left to get impact of error propagation.
  - sample 2-right with 3-left to get impact of imperfect concealment.
  - Now the sample content it the same. PESQ has not problems…
  - Calculate both times the importance values.
- Problem:
  - Split and merge introduces a “click” which falsified the results
  - Approach failed again!

Crossfading

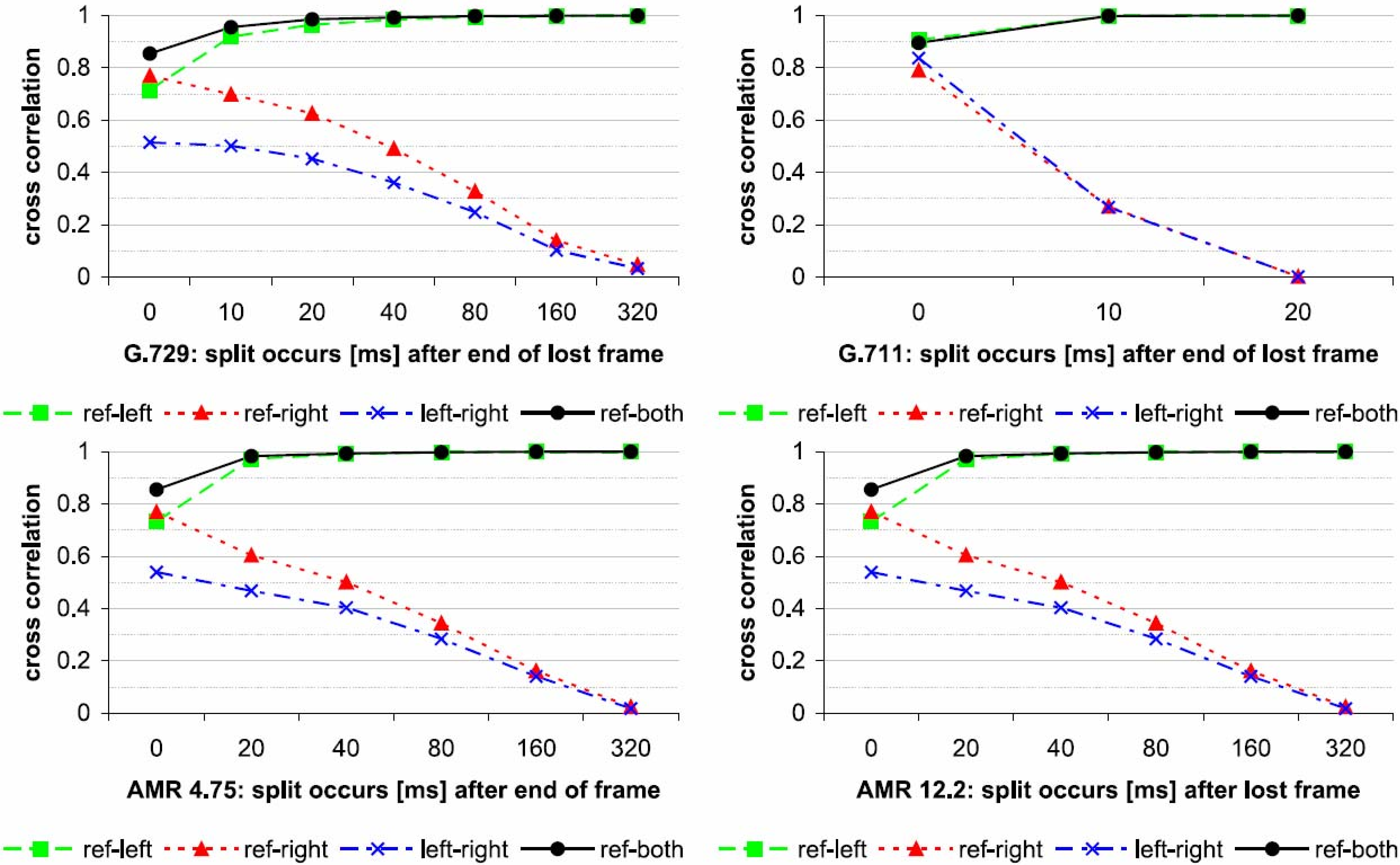


- Do not use hard split but cross-fading function (sinus curve)
- Cross fading length of 4 ms has proven to be a good compromise
  - Tradeoff between negative effect of the click and resolution in time.
- Additional Experiment:
  - Splitting can also occur somewhat after the frame loss.
  - To measure the time line of error propagation.

Measurement Results of Four Frame Losses during voice activity with a mean importance for four different codecs



Correlation between Importance Values
(considering different loss position, sample, speakers, and languages)



Conclusion

- Predicting the importance in real-time is difficult because
  - Loss impact depends on the amount of error propagation
  - EP is not known at the time of transmission.

- Include the next frames to predict the amount of error propagation
- Then, the importance calculation can be enhanced significant.
- Drawback: Increased algorithmic delay
- Good comprise:
  - Consider only 20-40 ms after the lost frame
  - to minimize false prediction due to error propagation effects.