Calculation of Speech Quality by Aggregating the Impacts of Individual Frame Losses

Christian Hoene

18th NMRG Meeting in Nancy, France

30. July 2005

Technische Universität Berlin

TKN Telecommunication Networks Group

URL: www.tkn.tu-berlin.de

Content

Based on a IWQoS'05 paper with the same title.

- Introduction
- Background
  - PESQ
  - Importance of Speech Frames
- Packet Dropping Strategy
- Adding
- Validation
- Summary

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 for the following frames.)
  - the surrounding speech.
- Example: Discontinuous Transmission (DTX)
  - Speech frames during silence are less important
  - Lower frame rate during silence

Prerequisite

The „Importance“ of a VoIP packet is understood:

- The speech packet’s importance is the quality degradation that its loss would cause.
- The importance of a speech frame can be measured with a verified approach.
- The importance values differ largely.
  - Some frames are very important
  - Others, even during voice activity, are negligible.

Goal

If we know the importance of ONE frame,

- how does **one** loss impact relate to **multiple** loss impacts?
- Or, how to calculate speech quality by **adding** the **importance** values of individual frame losses?

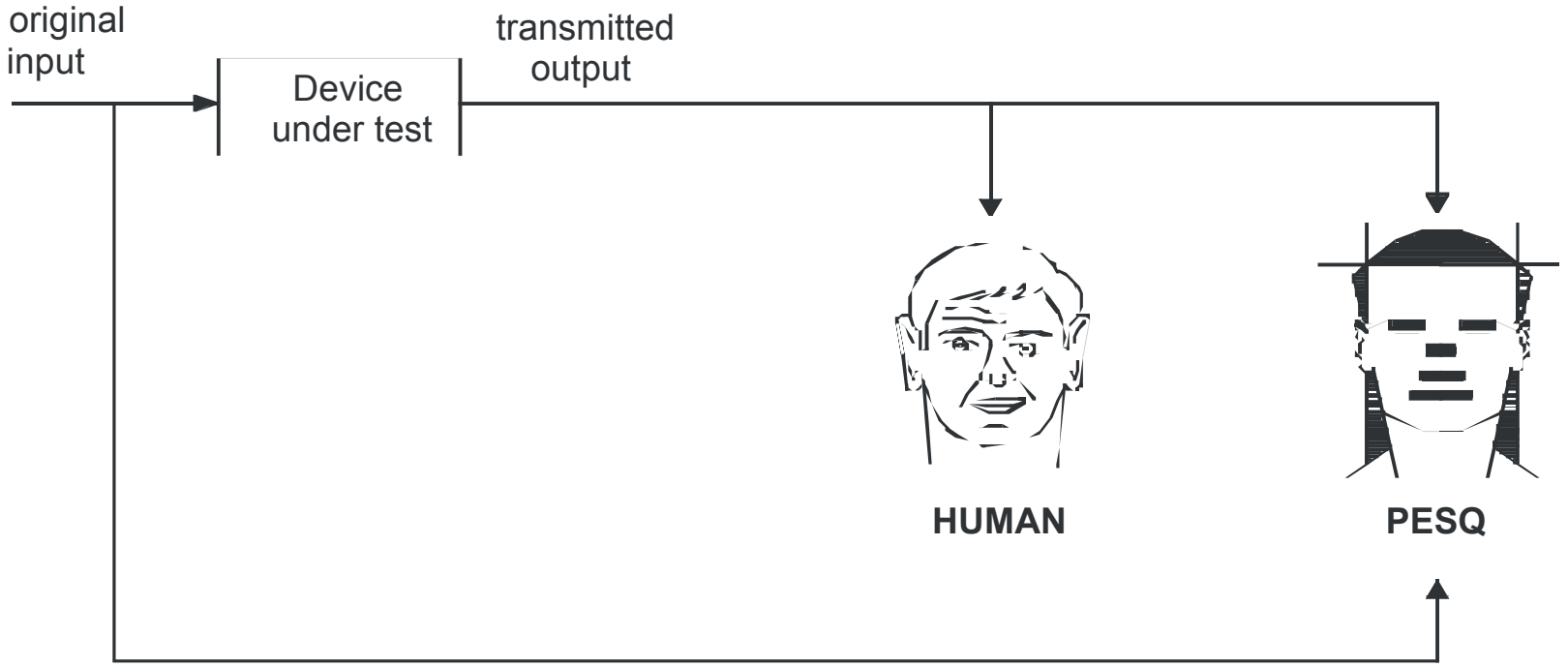
Content

- Introduction
- Background
  - PESQ
  - Importance of Speech Frames
- Packet Dropping Strategy
- Adding
- Validation
- Summary

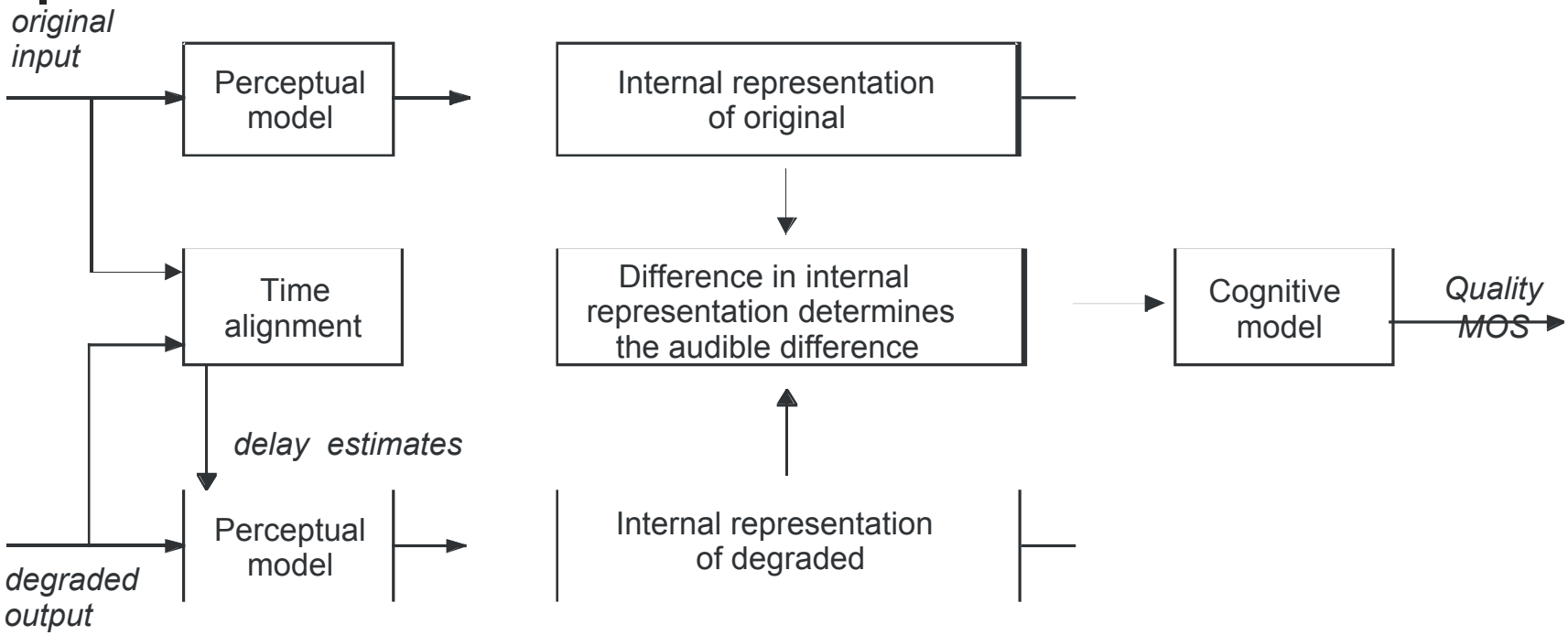
PESQ – Measuring Speech Quality.

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)

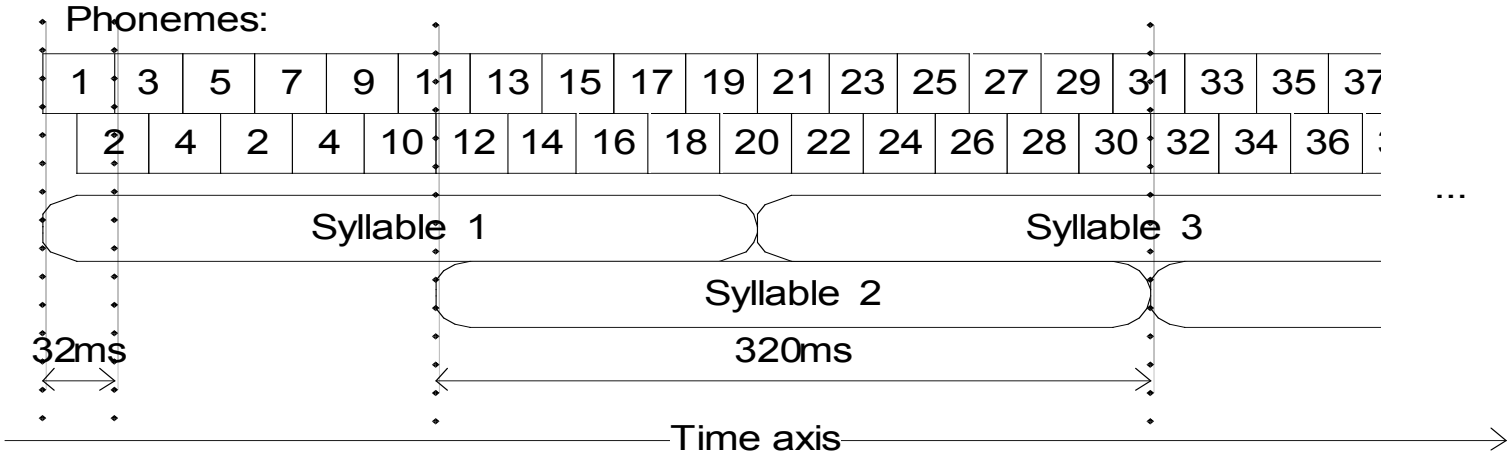


PESQ – Architectural Structure



- Time alignment to cope with transmission jitter.
- Perceptual modeling of speech signals.
- Comparison between original and degraded sample.

PESQ – Temporal Partitioning of Signal



- Signal is split to phonemes, which are 32 ms long.
- Twenty phonemes are summed up to one syllable.
- Phonemes as well as syllables are 50% overlapping!
  - Values have been chosen experimentally for high prediction accuracy.

Weighting of Disturbance over Time

- Calculate **disturbance** for each phoneme using asymmetric (A) and normal (D) perceptual difference.
- Sum up **phonemes** to **syllables** using Equation 2.
- Calculate overall disturbance using Equation 3.
- Get **$PESQ_{MOS}$** value with Equation 1.

$$PESQ_{MOS} = 4.5 - 0.1 \cdot D_{indicator} - 0.0309 \cdot A_{indicator} \quad (1)$$

$$syllable^{AorD}_{indicator}[i] = \sqrt[6]{\frac{1}{20}\sum_{m=1}^{20} phoneme^{AorD}_{disturbance}[m+10i]^6} \quad (2)$$

$$AorD_{indicator} = \sqrt{\frac{1}{N}\sum_{n=1}^{N} syllable^{AorD}_{indicator}[n]^2} \quad (3)$$

Content

- Introduction
- Background
  - PESQ
  - Importance of Speech Frames
- Packet Dropping Strategy
- Adding
- Validation
- Summary

Definition: Old Metric for Importance

The packet's importance is the quality degradation that its loss would cause.

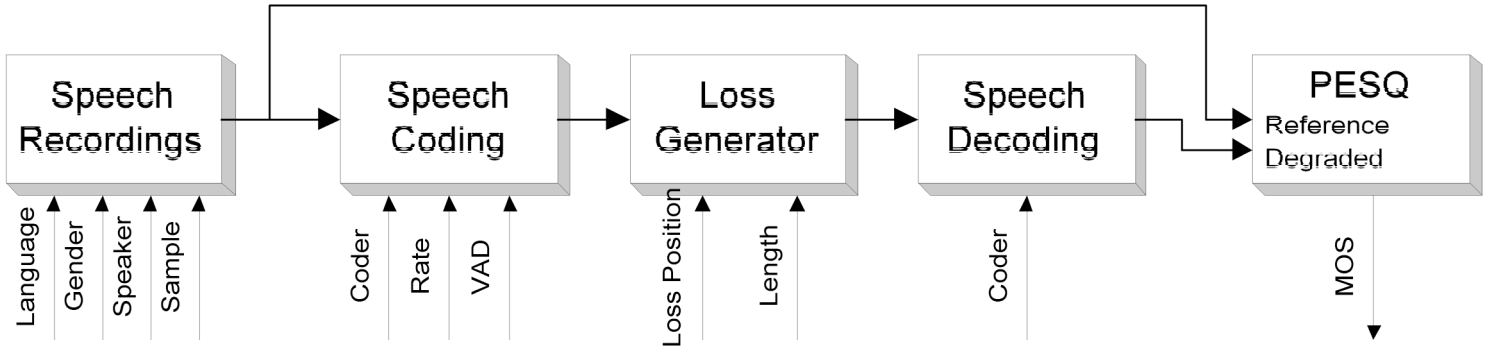
Definition:

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:

$$\mathrm{Imp}(s,c,e) = (\mathrm{MOS}(s,c) - \mathrm{MOS}(s,c,e)) \cdot t(s) \qquad (4)$$

s: sample
t(s): samples length (s)
c: codec implementation
e: loss event, one or multiple correlated frame losses

Measurement Setup (1)



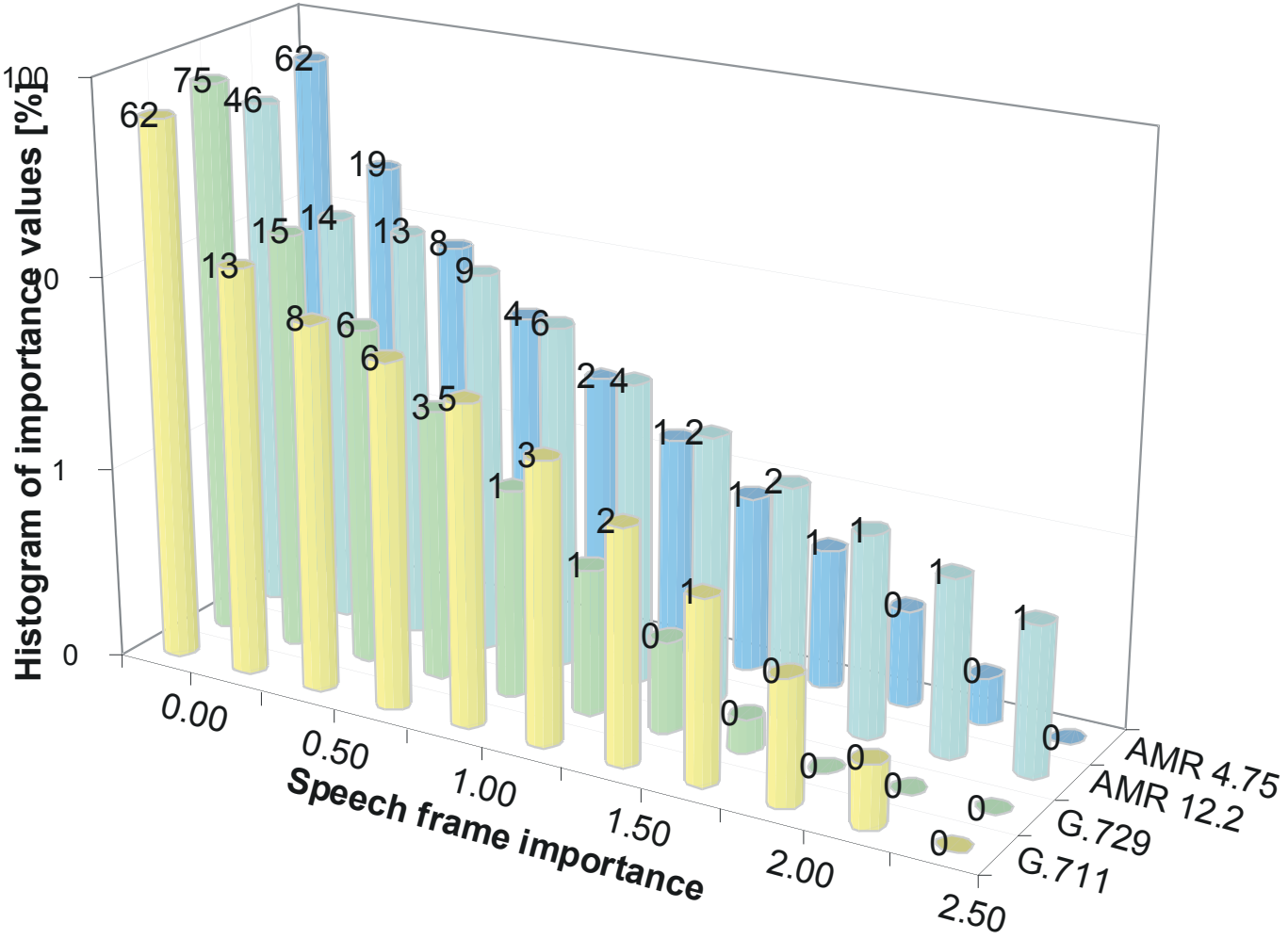
- Large sample database (ITU P suppl. 23)
  - 4 Languages x 4 speakers x 52 samples = 832
  - 8s each, two sentences
- Codec's:
  - ITU G.711 + Appendix II (64 kbit/s)
  - ITU G.729 (8 kbit/s)
  - 3GPP Adaptive Multi-Rate (4.75...12.2 kbit/s)

Measuring the Importance of a VoIP frame

- Using PESQ to measure the loss of **one** speech frame
- Speech frames differ largely

Mean Importance [MOS*s]

	AMR 4.75	G.729	AMR 12.2	G.711
active	0.389	0.655	0.923	1.338
silence	0.003	0.004	0.008	0.016



PESQ not verified ⇨ conduct listening tests

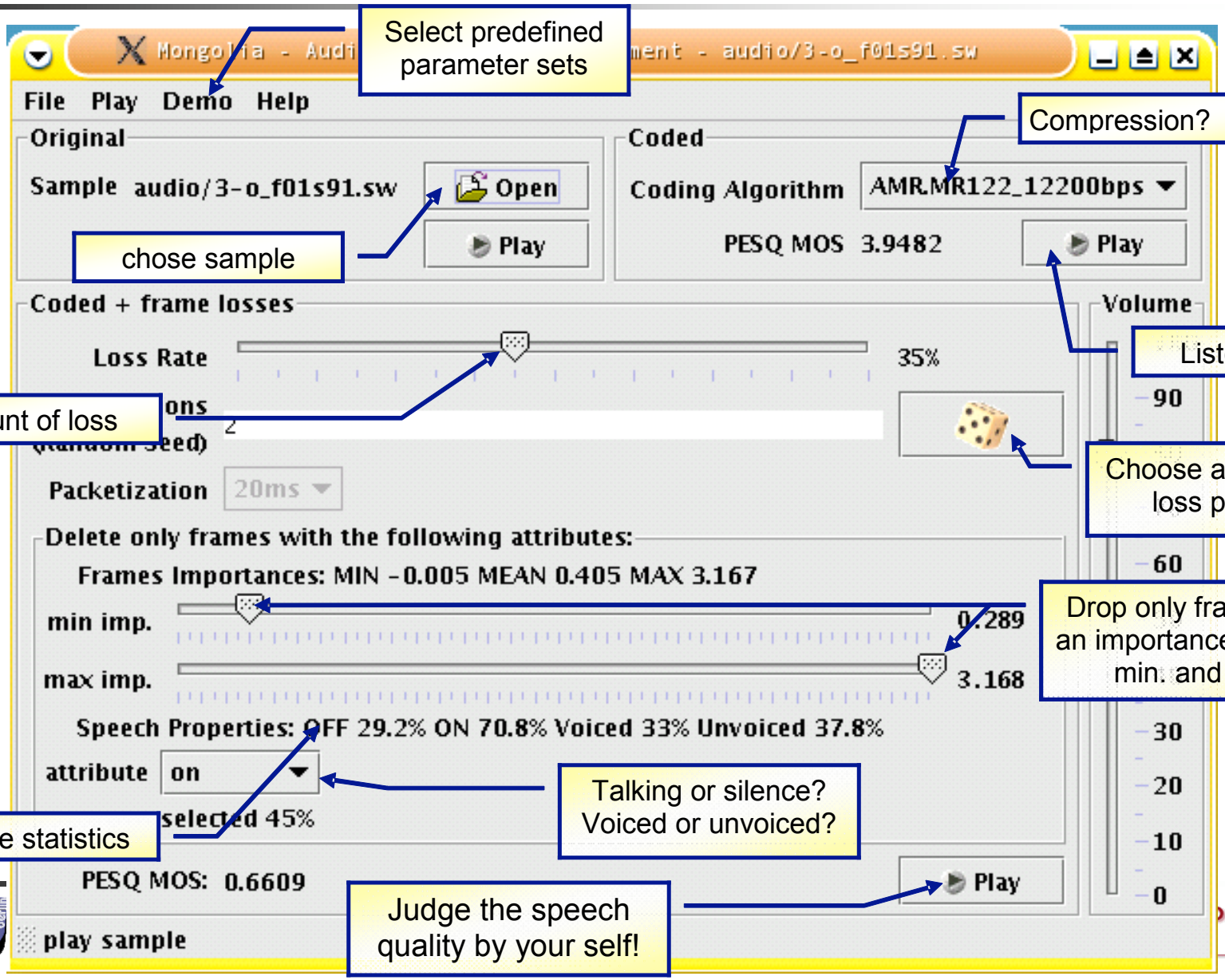
- Can PESQ measure frame importances?
  Nobody knows…

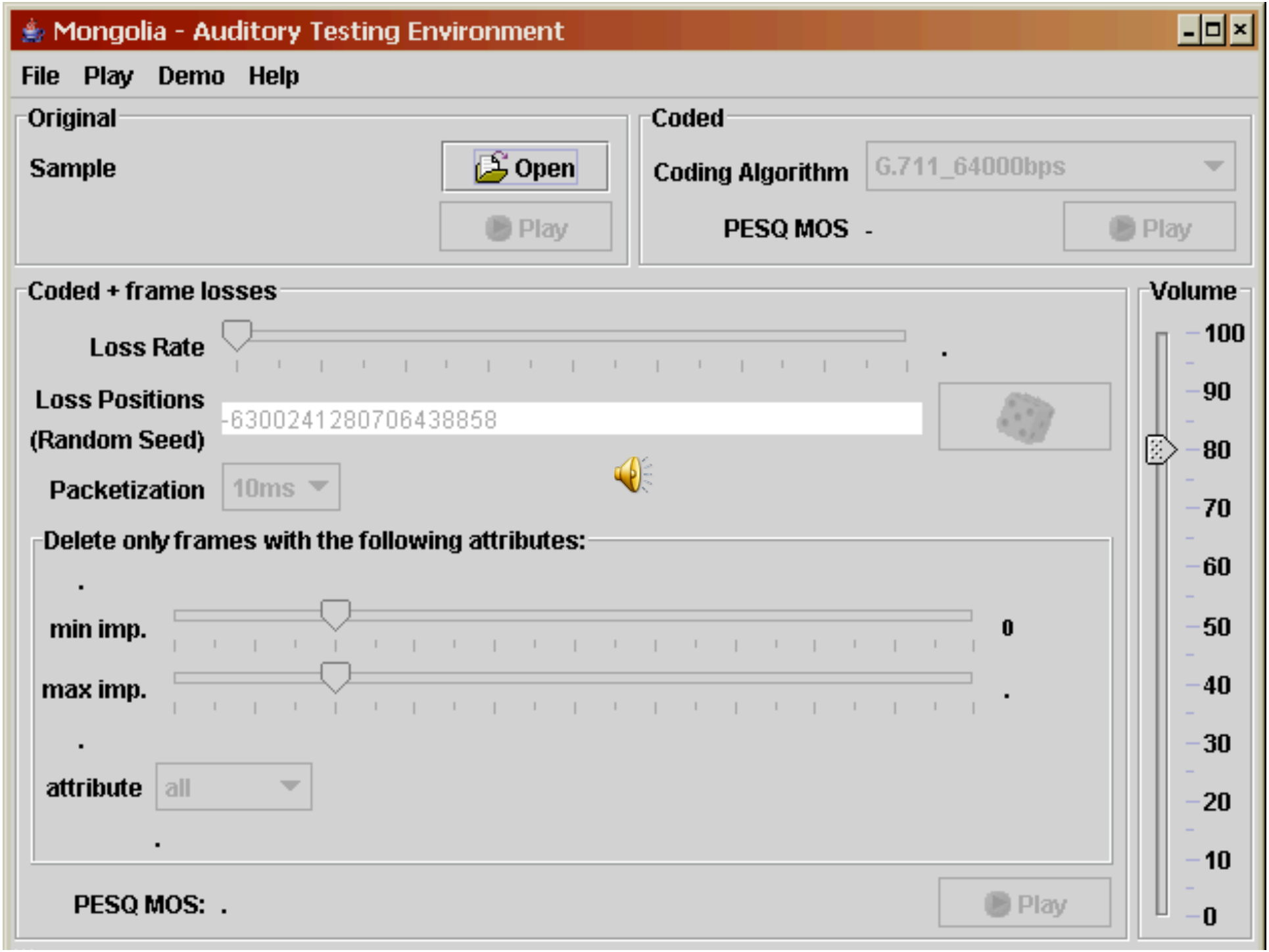
⇨Thus: Conduct formal listening-only tests!

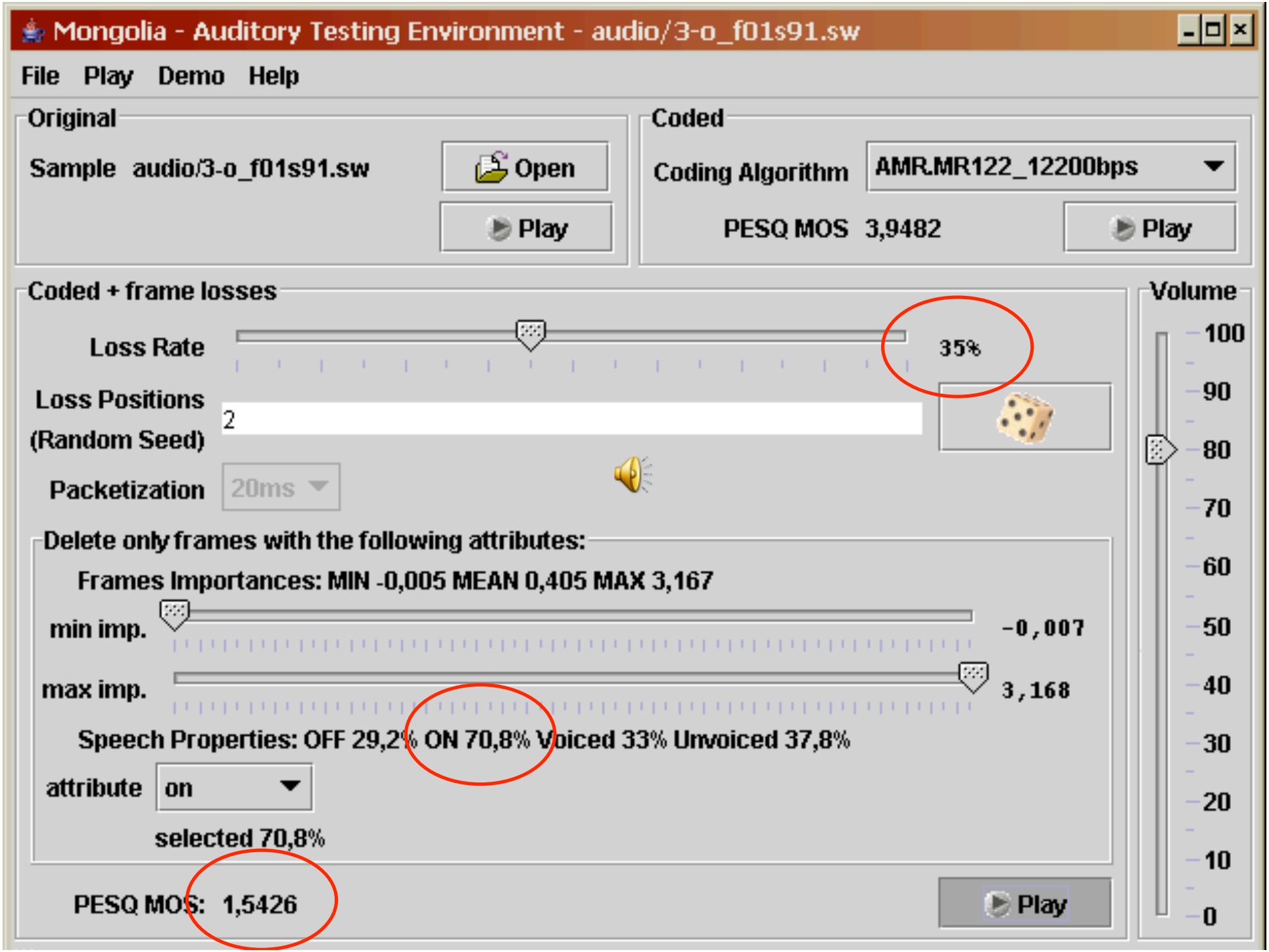
Problem:
Humans cannot hear single frame losses!

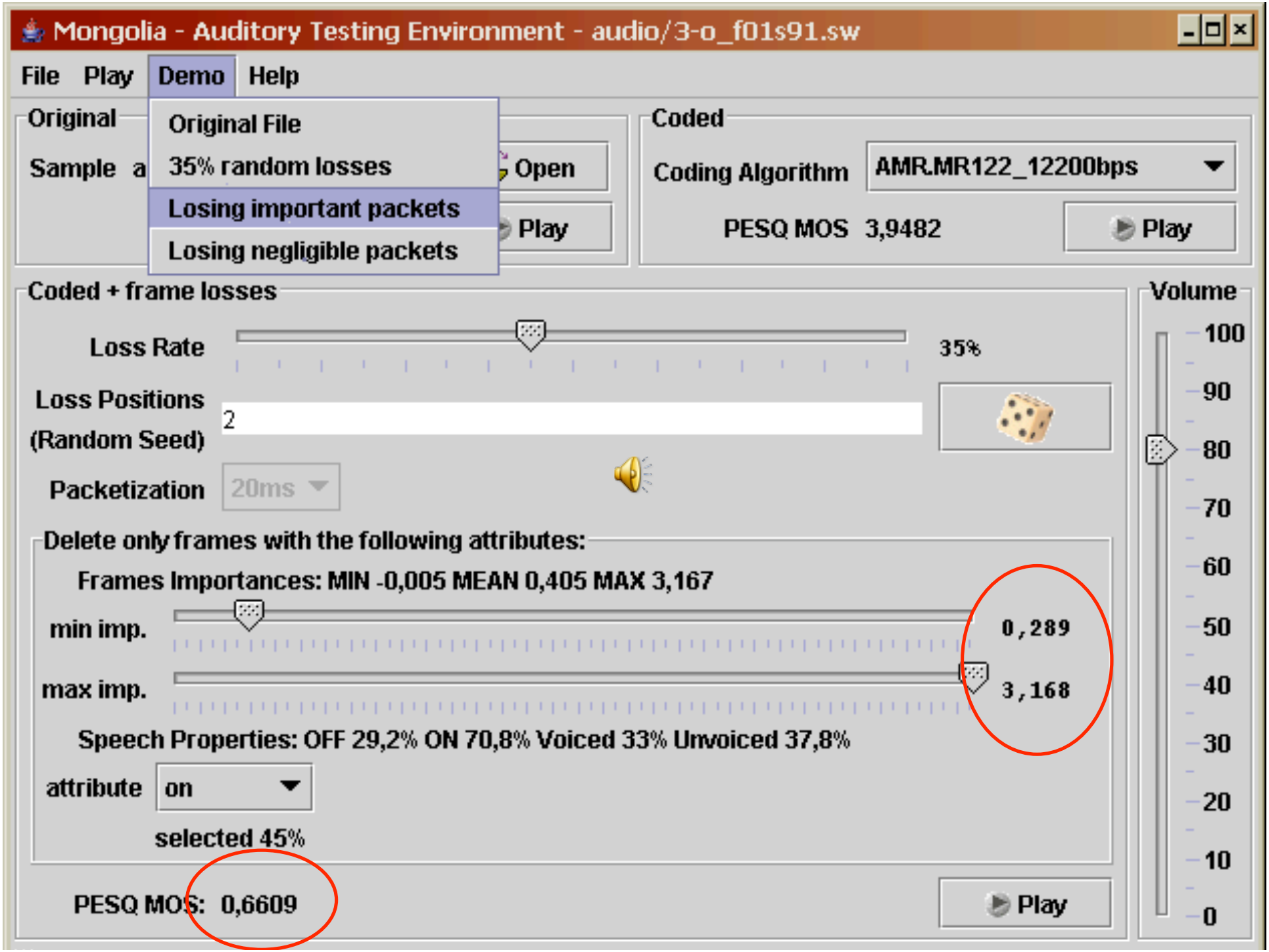
- Human just can hear multiple frame losses.
- Thus, drop multiple similar frames…
- PESQ can identify similar frames.

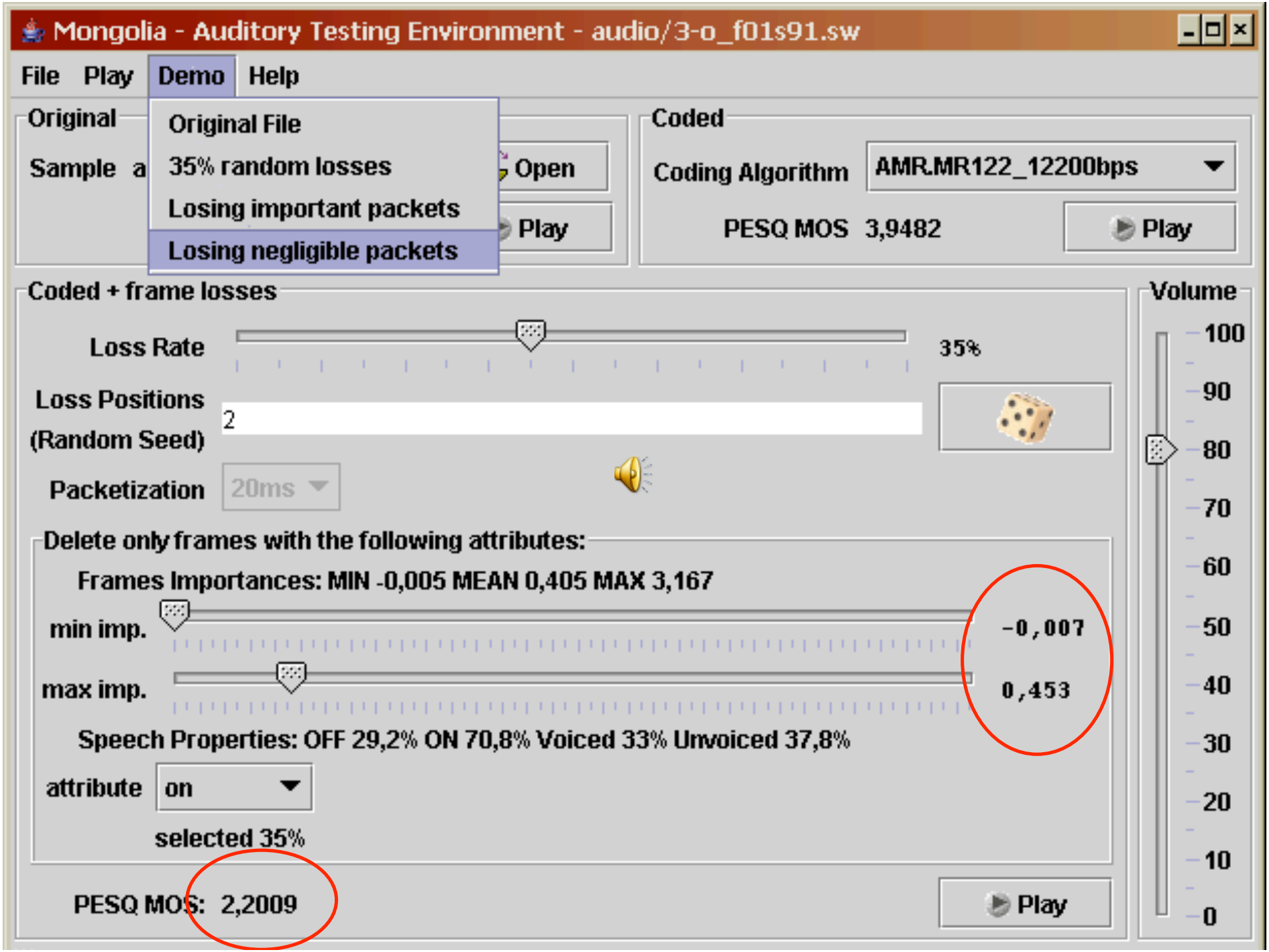
Just try it: www.tkn.tu-berlin.de/research/mongolia



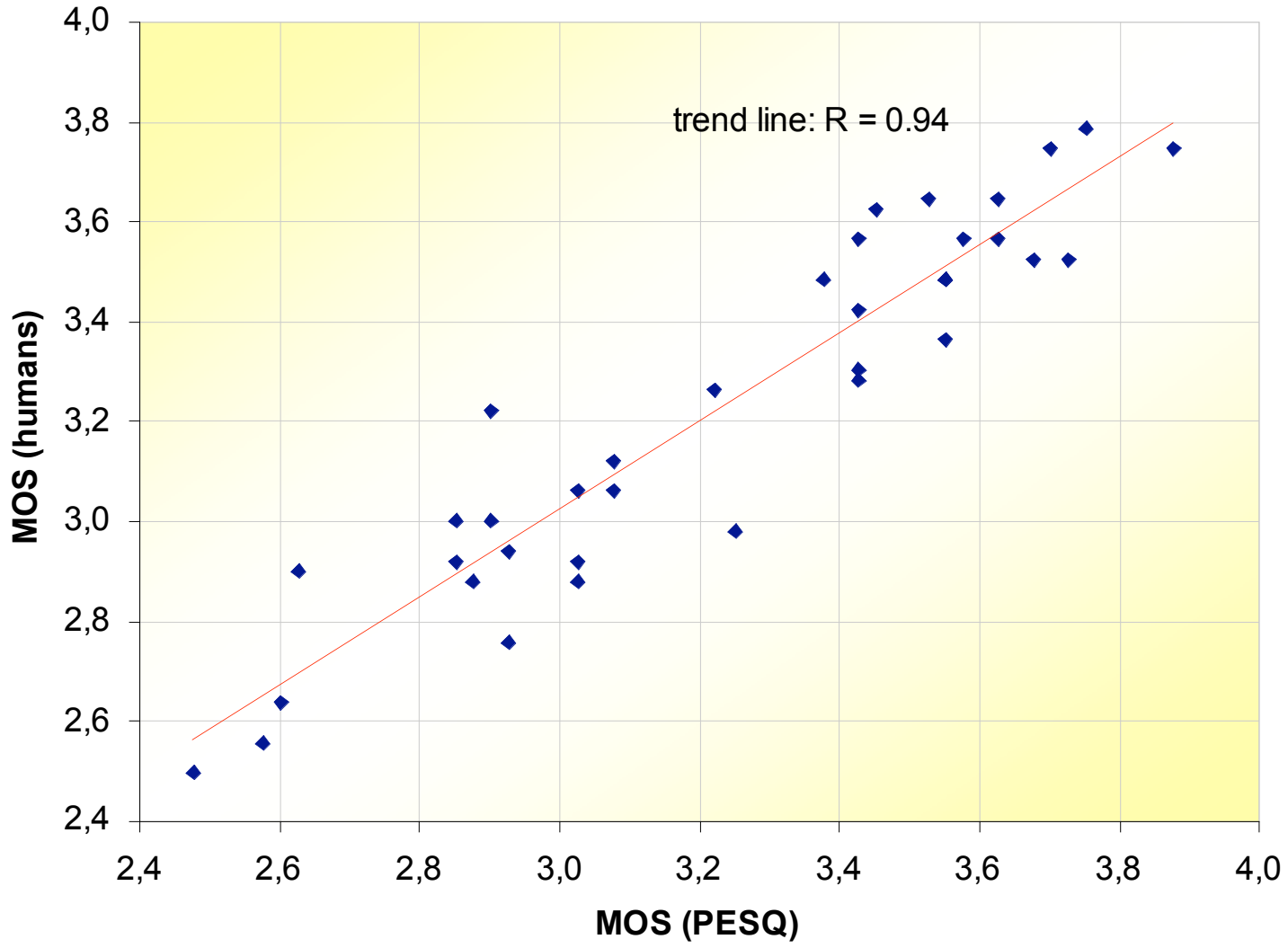

Mongolia - Auditory Testing Environment
File
Play
Demo
Help
Original
Sample
Open
Play
Coded
Coding Algorithm
G.711_64000bps
PESQ MOS -
Play
Coded + frame losses
Loss Rate
Loss Positions
(Random Seed)
-6300241280706438858
Packetization
10ms
Delete only frames with the following attributes:
min imp.
0
max imp.
attribute
all
PESQ MOS: .
Play
Volume
100
90
80
70
60
50
40
30
20
10
0


Mongolia - Auditory Testing Environment - audio/3-o_f01s91.sw
File
Play
Demo
Help
Original
Sample audio/3-o_f01s91.sw
Open
Play
Coded
Coding Algorithm
AMR.MR122_12200bps
PESQ MOS 3,9482
Play
Coded + frame losses
Loss Rate
35%
Loss Positions
(Random Seed)
2
Packetization
20ms
Delete only frames with the following attributes:
Frames Importances: MIN -0,005 MEAN 0,405 MAX 3,167
min imp.
-0,007
max imp.
3,168
Speech Properties: OFF 29,2% ON 70,8% Voiced 33% Unvoiced 37,8%
attribute
on
selected 70,8%
PESQ MOS: 1,5426
Play
Volume
100
90
80
70
60
50
40
30
20
10
0


Mongolia - Auditory Testing Environment - audio/3-o_f01s91.sw
File
Play
Demo
Help
Original
Original File
35% random losses
Losing important packets
Losing negligible packets
Sample a
Open
Play
Coded
Coding Algorithm
AMR.MR122_12200bps
PESQ MOS 3,9482
Play
Coded + frame losses
Loss Rate
35%
Loss Positions
(Random Seed)
2
Packetization
20ms
Delete only frames with the following attributes:
Frames Importances: MIN -0,005 MEAN 0,405 MAX 3,167
min imp.
0,289
max imp.
3,168
Speech Properties: OFF 29,2% ON 70,8% Voiced 33% Unvoiced 37,8%
attribute
on
selected 45%
PESQ MOS: 0,6609
Play
Volume
100
90
80
70
60
50
40
30
20
10
0


Mongolia - Auditory Testing Environment - audio/3-o_f01s91.sw
File
Play
Demo
Help
Original
Original File
35% random losses
Losing important packets
Losing negligible packets
Sample a
Open
Play
Coded
Coding Algorithm
AMR.MR122_12200bps
PESQ MOS 3,9482
Play
Coded + frame losses
Loss Rate
35%
Loss Positions
(Random Seed)
2
Packetization
20ms
Delete only frames with the following attributes:
Frames Importances: MIN -0,005 MEAN 0,405 MAX 3,167
min imp.
-0,007
max imp.
0,453
Speech Properties: OFF 29,2% ON 70,8% Voiced 33% Unvoiced 37,8%
attribute
on
selected 35%
PESQ MOS: 2,2009
Play
Volume
100
90
80
70
60
50
40
30
20
10
0

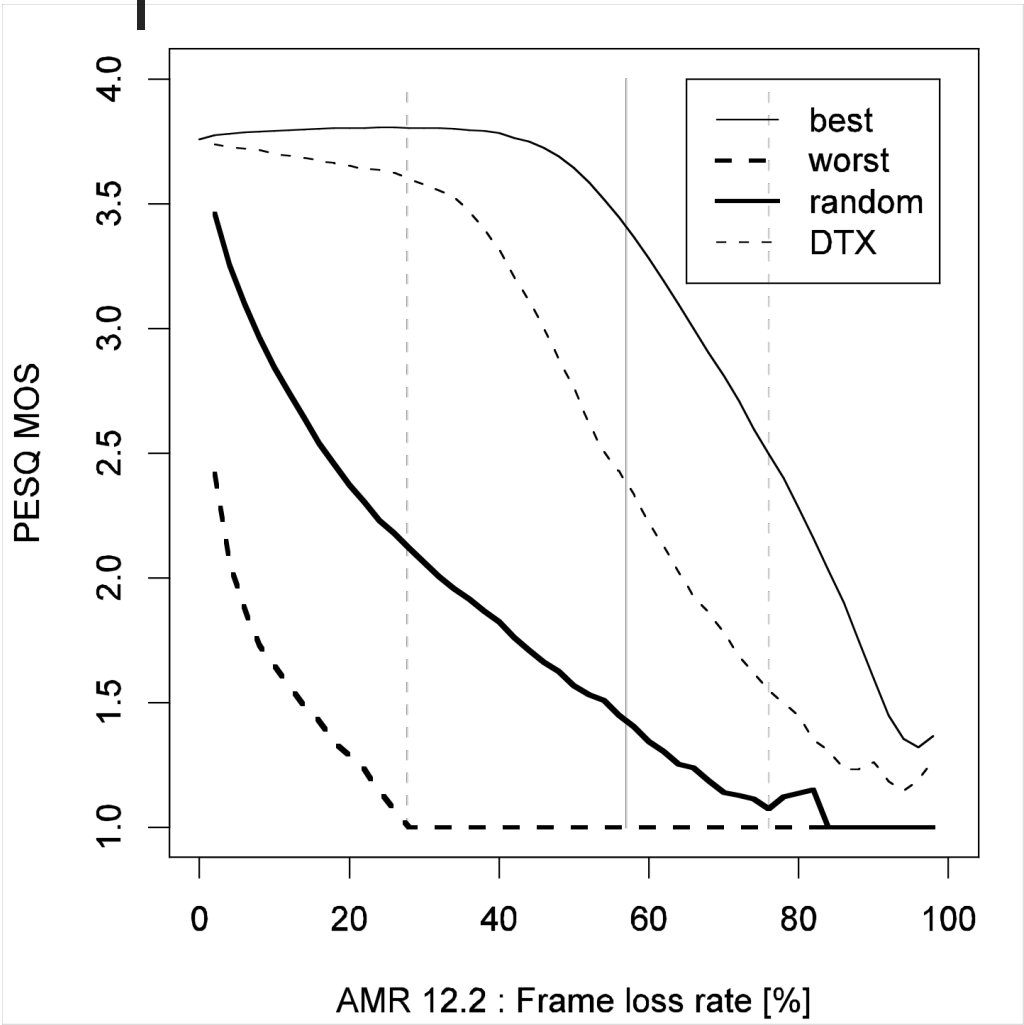
Human LQS-MOS vs. PESQ LQO-MOS



Content

- Introduction
- Background
  - PESQ
  - Importance of Speech Frames
- Packet Dropping Strategy
- Adding
- Validation
- Summary

Speech Frames – Dropping Strategy [8]

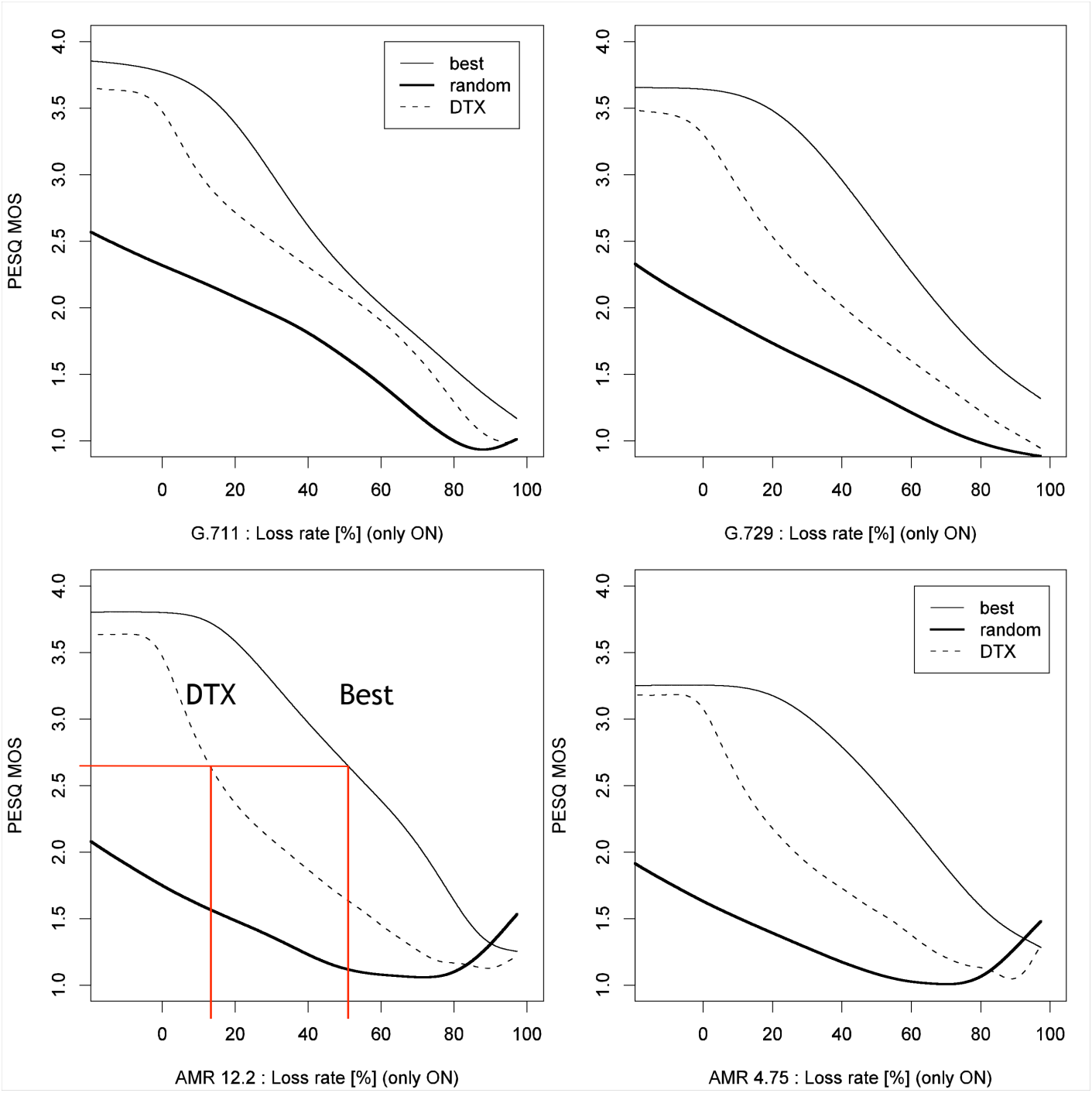


Lost packets in cases of

- Congestion
- Wireless fading
- Saving energy

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

Counting only active (ON) speech frames.



Results

- Speech frames and VoIP packets differ greatly.
- If packets have to be dropped, drop unimportant packets first.
- An offline method to measure the importance has been developed and verified.
- A real-time algorithm is required for telephony (future publication).

Content

- Introduction
- Background
  - PESQ
  - Importance of Speech Frames
- Packet Dropping Strategy
- Adding
- Validation
- Summary

Requirements on a Importance Metric

- Easy to calculate and measure (e.g. using PESQ).
- One-dimensional for simple use.
- It should be possible to give a statement like:
  - FrameA and FrameB are as important as FrameC, or
  - FrameD is three times more important than FrameE.
  - This is called **additive property**.

- Required for analytical models,
  e.g. for **rate-distortion** (RaDiO) optimized multimedia streaming **by Chou** and Miao

Approach

- Remodeling the behavior of PESQ for frame losses:
- Using a similar algorithm for aggregate frame losses as PESQ uses for speech signals.

- We develop a scale, which allows to **ADD** linearly frame importance values.
  - Works well for distant losses.

- Sorry, I will skip the analytical explanation, showing only the results:

$$Imp\left(s,c,e\right) = \left(cl - c\right) \cdot t\left(s\right)$$
$$\text{with } cl = \left(4.5 - MOS\left(s,c,e\right)\right)^2 \text{ and } c = \left(4.5 - MOS\left(s,c\right)\right)^2 \qquad (10)$$

Short-term aggregation

- We show that if frame losses occur shortly one after the other, temporal auditory masking effects have to be considered.
  - A heuristic equation to model these effects:

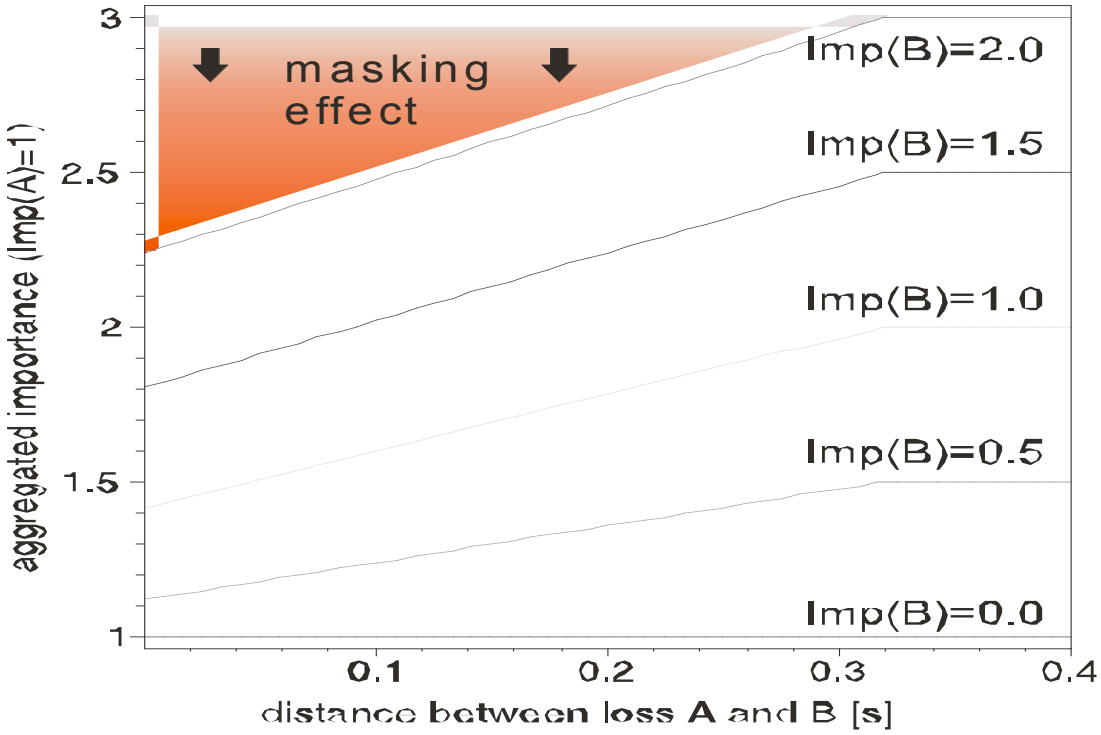
- Probabilitv that two losses occur in the same:

$$P_{in.syll}\left(t_{width}\right) = \frac{1}{t_{syll}} \int_{t_a=0}^{t_{syll}} \left\{ \begin{array}{ll} 0 \text{ if } t_a + t_{width} \geq t_{syll} \\ 1 \qquad \text{otherwise} \end{array} \right\} \mathrm{d}t_a$$
$$= \left\{ \begin{array}{ll} 0 & \text{if } t_{width} \geq t_{syll} \\ 1 - \frac{t_{width}}{t_{syll}} & \text{otherwise} \end{array} \right. \qquad (11)$$

t_{width} distance between both losses

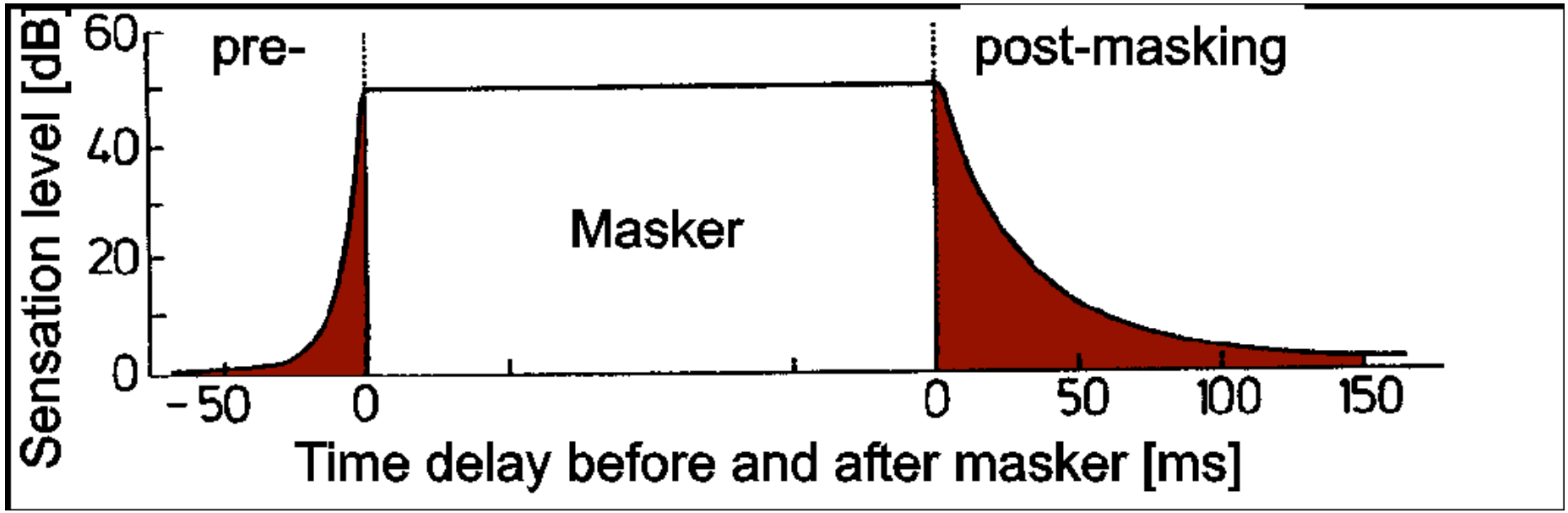
t_{syll} length of syllable (320ms)

Aggregated Loss Impact



- The aggregated loss impact depend on the individual importance values,
- and the **distance between losses**!
- Actually, a **masking effect** is modeled!

Psychoacoustic masking effects

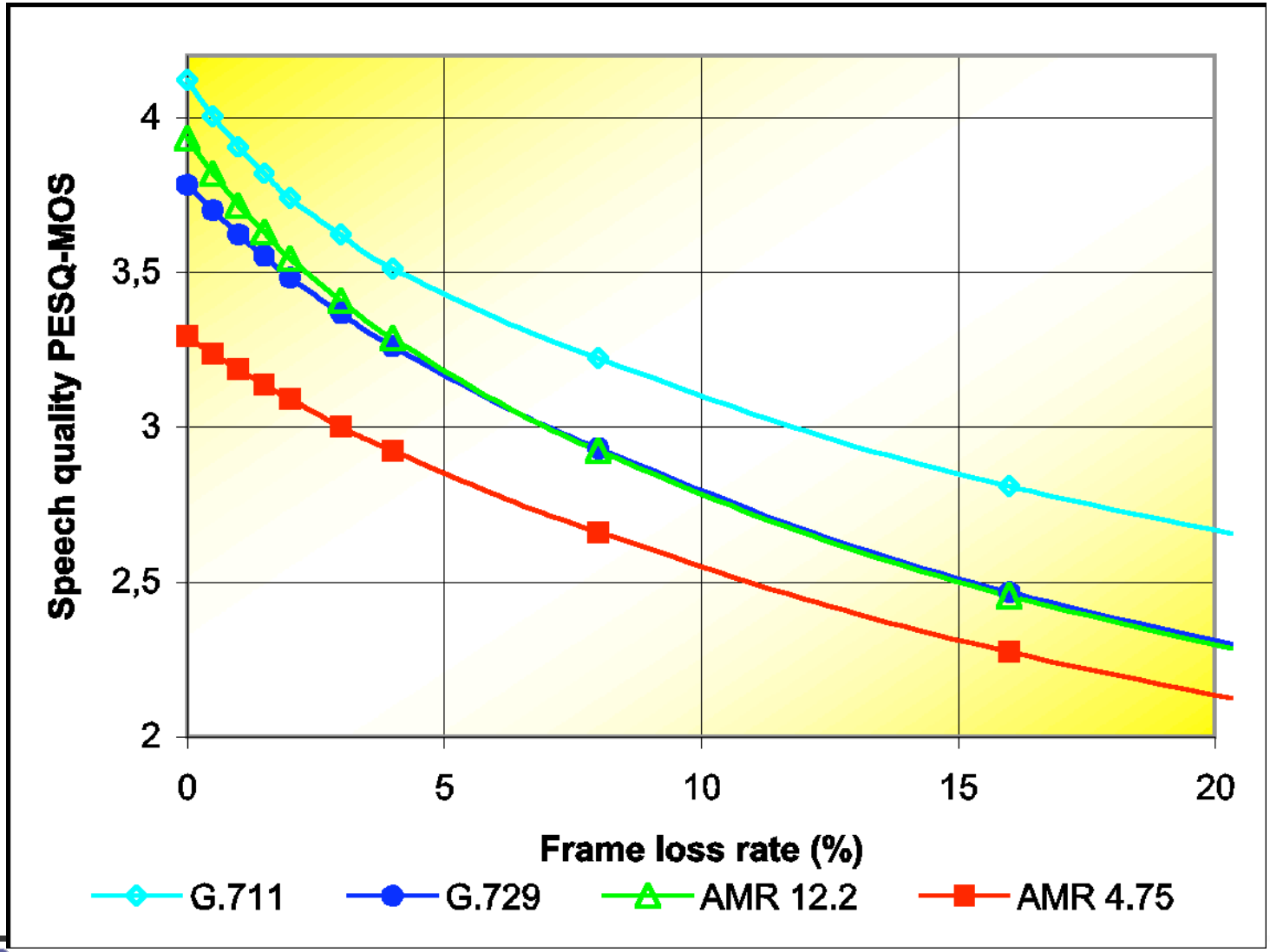


- Masking effect are not included directly in PESQ.
- However, PESQ's weighting-over-time models a similar effect indirectly.

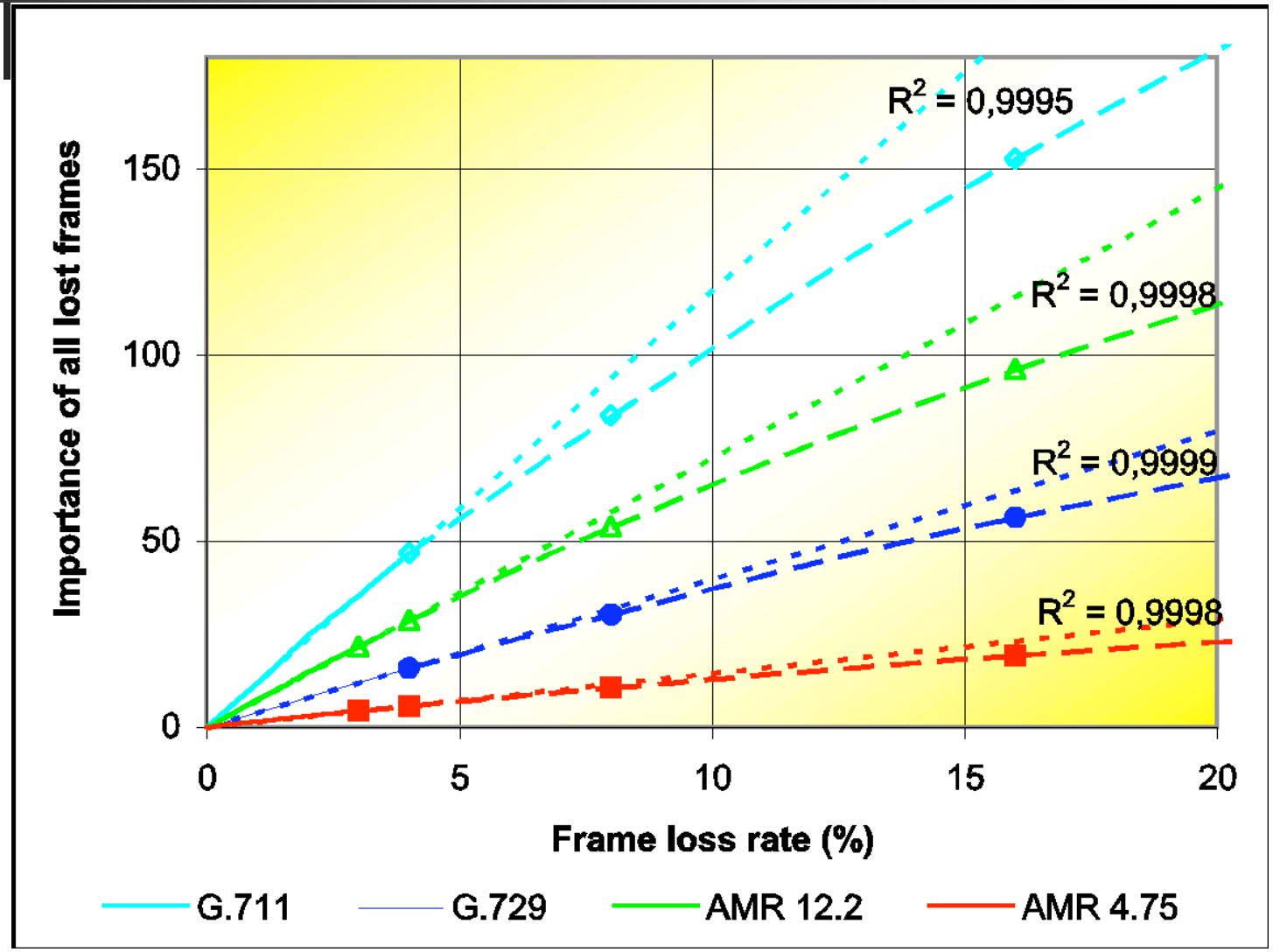
Content

- Introduction
- Background
  - PESQ
  - Importance of Speech Frames
- Packet Dropping Strategy
- Adding
- Validation
- Summary

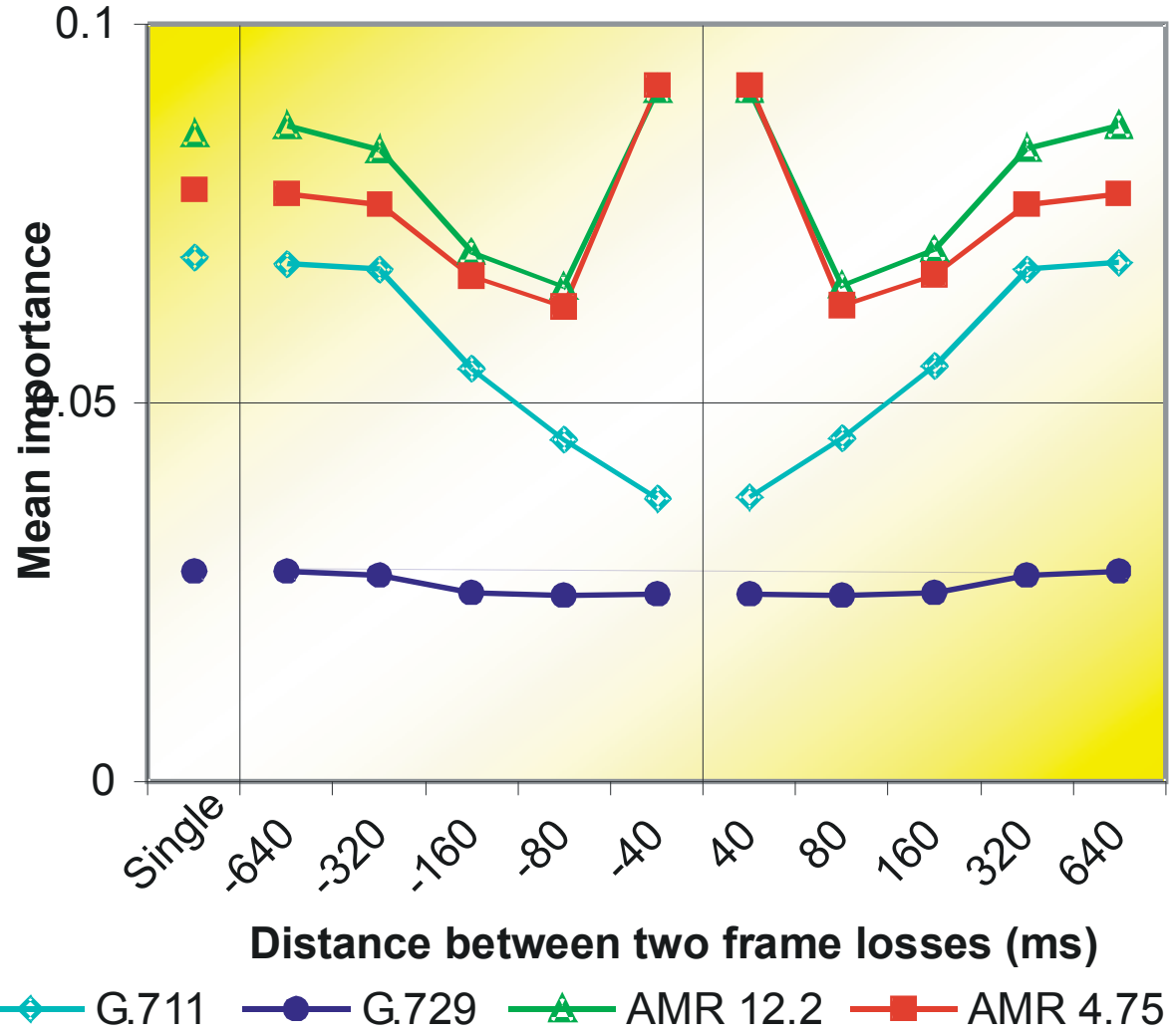
Impact of random losses on speech quality.



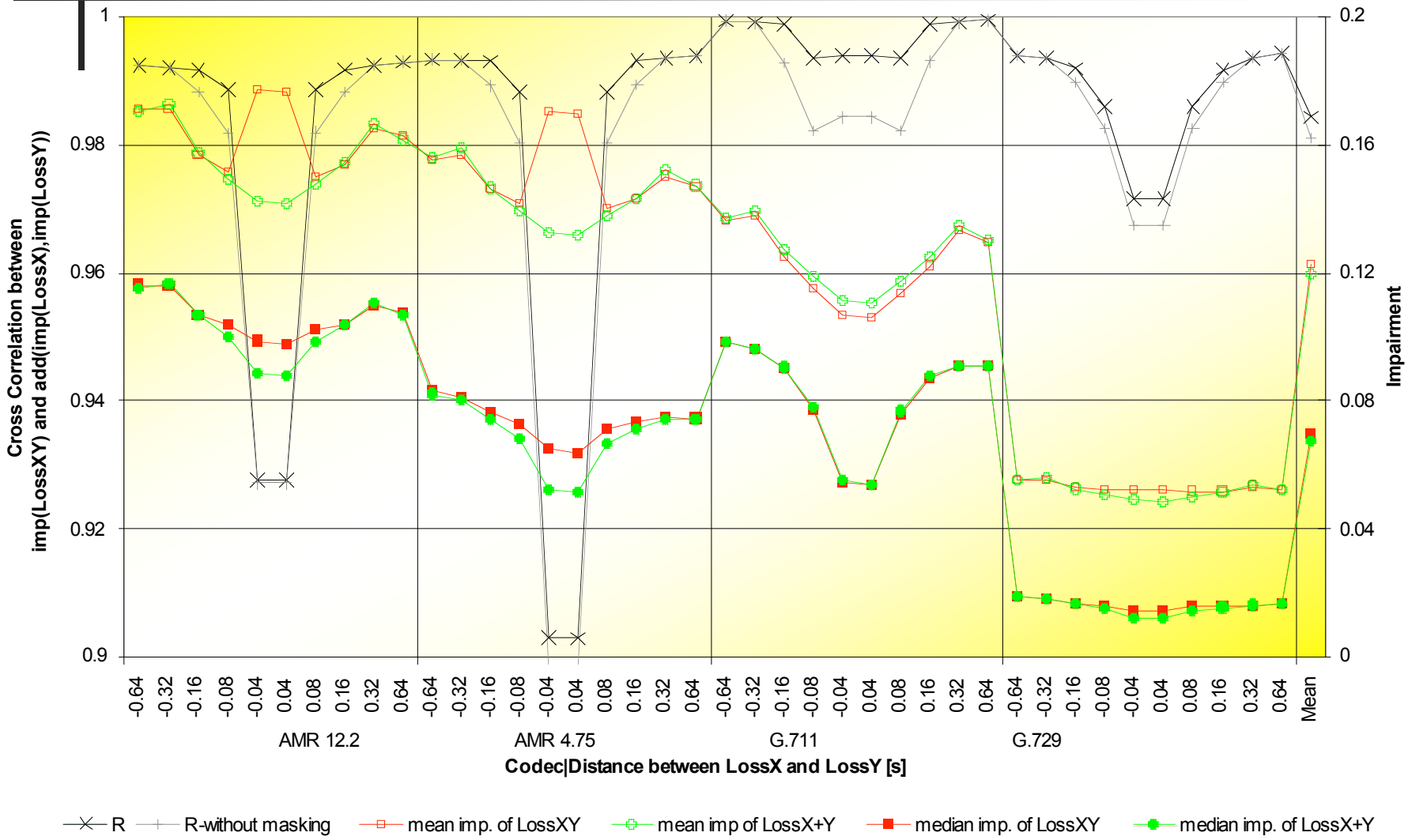
Impact on random losses on importance



Importance of second loss, depending on distance between two losses



Correlation coefficiency (R) between our model and PESQ.



Limits of our model (to do list)

- The effect of error propagation is not modeled, yet.
- The effect of concealment in cases of bursty losses needs to be considered.
- Verification with subjective (human based) speech ratings.

- Can be used standardization in ITU-T P.VTQ or ITU-T G.107.

Content

- Introduction
- Background
  - PESQ
  - Importance of Speech Frames
- Packet Dropping Strategy
- Adding
- Validation
- Summary

Summary

- Substantial reduction of energy consumption
- if only important frames are transmitted (e.g. for Wifi VoiP phones?).
- We presented a new metric to describe the importance of a speech frame
- and an aggregation function considering post-masking effects.

WWW: http://www.tkn.tu-berlin.de/~hoene/

L'Excelsior

50, rue Henri-Poincaré
54000 Nancy

Phone : 33 (0)3 83 35 24 57
Fax : 33 (0)3 83 35 18 48

RESERVE YOUR TABLE

Breakfast from 8am to 11.30am,
afternoon tea from 3pm to 6.30pm