VoIP on WIFI

Christian Hoene

September 13th, 2004

UNITN – DIT, Trento

Technische
Universität
Berlin

TKN Telecommunication Networks Group

URL: www.tkn.tu-berlin.de

Content

- Motivation: Semantic data-link
- Determining VoIP Quality
  - Perceptual quality of VoIP packet traces
  - Listening-only tests
- On the Importance of a VoIP Packet
  - Audio Demos
  - Predicting the Importance
- Voice over WLAN
  - Experiments: Semantic data-link
  - Simulations: Voice over EDCA
- Conclusions

Phone Calls on WLAN?

State of the Art::

Two technologies for voice-(DECT)

and data services (IEEE 802.11)

- Increased total cost of ownership (TCO) for 2x purchase, 2x operation, 2x administration
- Do customers prefer combined products?
  - Data over DECT?
  - Voice over WLAN?
  - A new wireless technology for voice and data?

⇨ Digital packet telephone services (with VoIP) on wireless local networks using IEEE 802.11+

Research Challenges

WLAN phones are often worse than DECT phones:

- Higher energy consumption
- Coverage is smaller
- Tel-systems are more complex and expensive.

⇨ Optimization of speech transmission over WLAN in order to achieve:

1. Better speech quality for a given capacity
2. Higher capacity at the same level of speech quality.
3. Reduction of energy consumption

How to Improve Speech Communication?

- New speech coding algorithms ("codecs")
- Classic approaches (e.g. in UMTS)
  - combined source channel coding
  - unequal bit error protection
  - soft-bit decoding
- Algorithms in packet networks
  - Separating different data flows (IntServ, DiffServ)
  - Overprovision (increase bandwidth)
  - Medium access protocols (PCF, 802.11e)
  - High priority for delay-sensitive flows

Enhancing Voice over WLAN

Where to enhance?

- Backbones are fast. They are not a problem.
- WLAN access is the bottleneck!

What to change?

- Codecs are given: ITU standards
- Transport/Network: IETF RFC standards
- Data-link and physical layer:
  IEEE 802.11... standards.

But: Change parameters sets on the fly!

Semantic Transmission

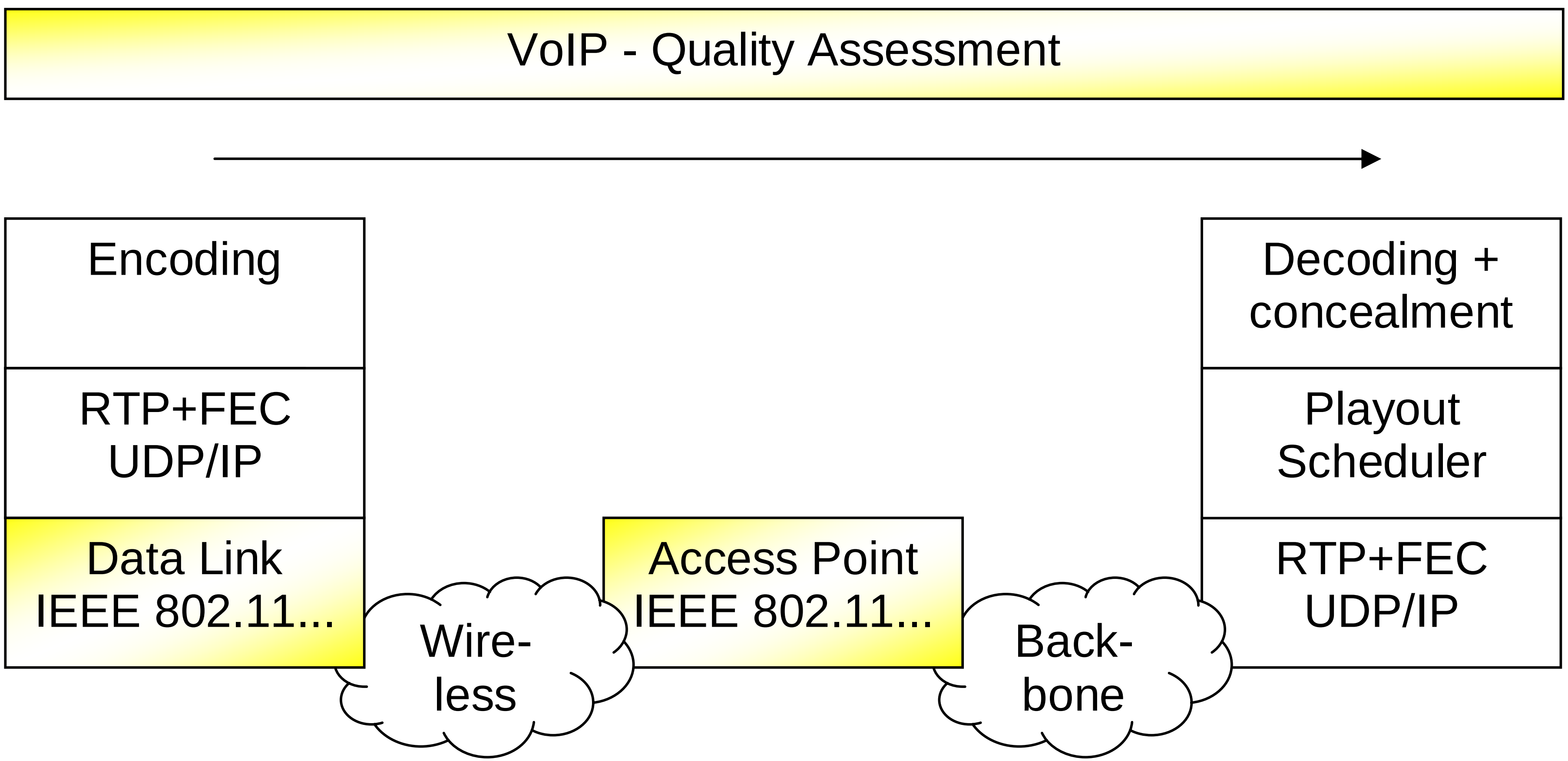
- Utilize the temporal variance of speech:
  - Of course: transmitted during voice activity and wait during silence
  - But: even during voice activity speech segments differ greatly.

⇨ *Transmit every speech frame according to its importance.*

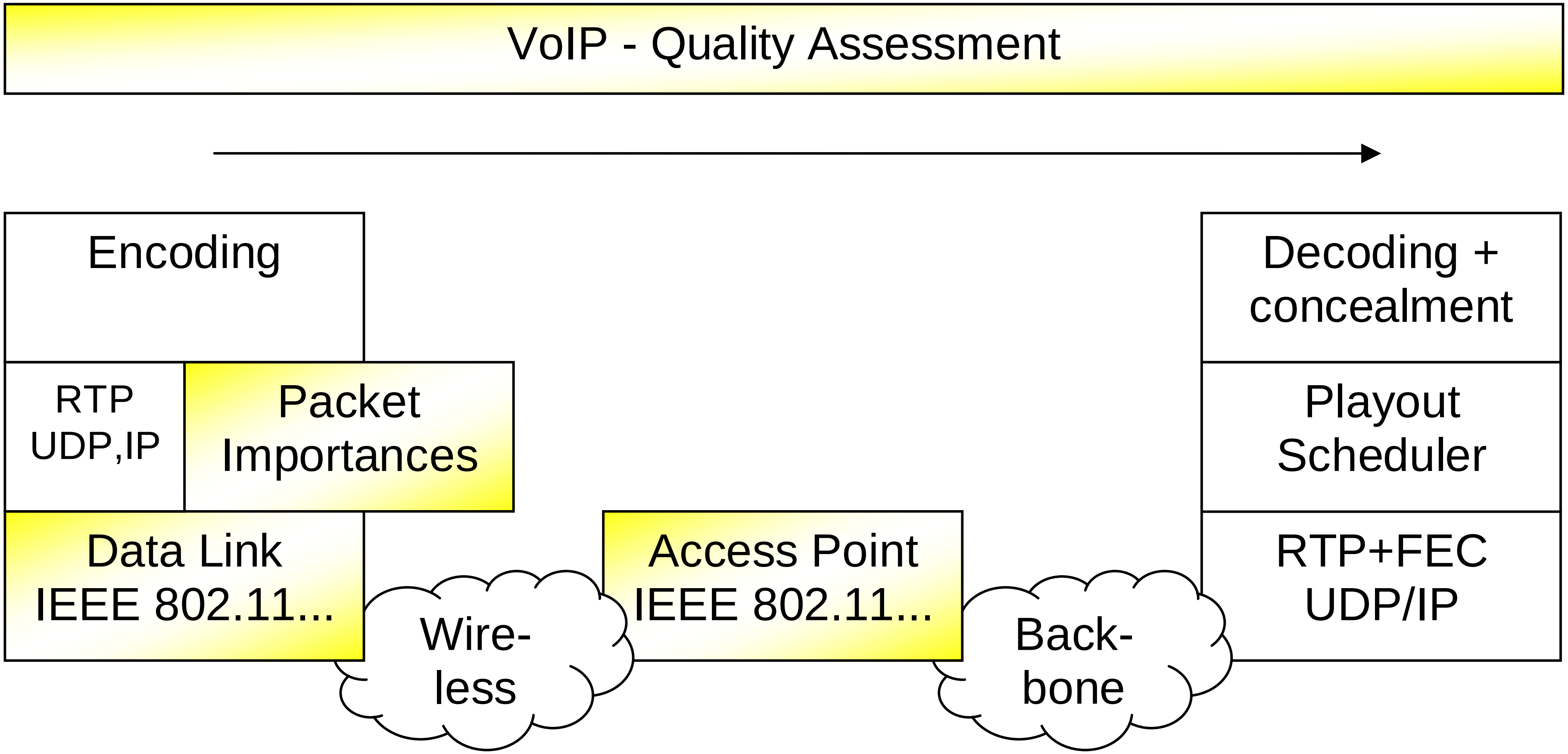
⇨ Drop the unimportant frames sooner than more important ones

(e.g. voiced sounds are more important than unvoiced sounds).

Classic Protocol Stack



Semantic Data-Link



Content

- Motivation: Semantic data-link
- Determining VoIP Quality
  - Perceptual quality of VoIP packet traces
  - Listening-only tests
- On the Importance of a VoIP Packet
  - Audio Demos
  - Predicting the Importance
- Voice over WLAN
  - Experiments: Semantic data-link
  - Simulations: Voice over EDCA
- Conclusions

How to Judge the Quality of a Call?

Measure the quality of a Wifi-Voice system:

- Gather packet loss rates, mean delay, etc.
  - 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)

Assessing a Trace of VoIP Packets

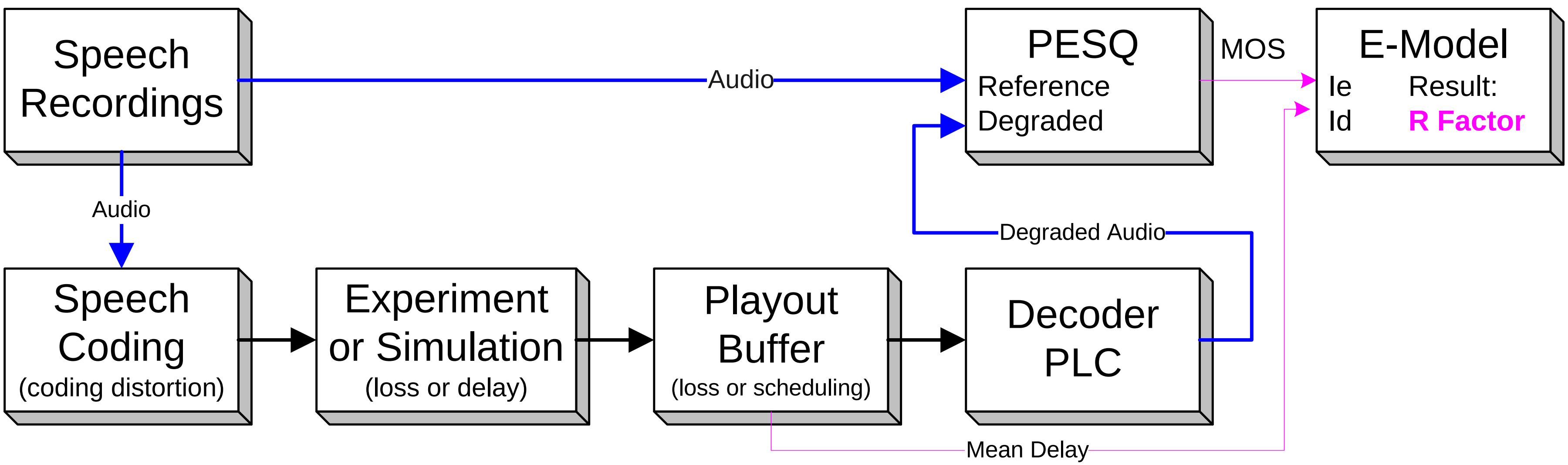
PESQ and E-Model cannot judge the impact of:

- Variable packet delay ⇨ adaptive playout time
- Packet losses are considered to be random
  but depend on the content and context of the speech frame
- Not applicable to assess VoIP packets traces
  resulting from simulations or experiments

Our solution:

- Using different playout schedulers and encoding schemes
- combine von E-Model und PESQ *(approach approved by the ITU)*
- PESQ calculate speech quality
- E-Model to combine MOS rating and transmission delays

Speech Quality and Transmission Delay



$$R_{factor} = \mathrm{MOS_2R}\left(\mathrm{MOS_{PESQ}}\right) - I_{dd}\left(t\right)$$

Equations – referring to the SPECTS'04 paper…

If neither talker nor listener echoes are present, the delay impairment I_d can be reduced to the term of I_{dd}: For an end-to-end delay $0 < T_a \leq 100\,\text{ms}$, I_{dd} is 0. For any $100\,\text{ms} \leq T_a < 500\,\text{ms}$ is

$$I_{dd}\left(T_a\right) = 25\left(\left(1 + X^6\right) - 3\left(1 + \left(\frac{X}{3}\right)^6\right)^{\frac{1}{6}} + 2\right) \tag{3}$$

with $X = \frac{-2+\lg T_a}{\lg 2}$. The mean opinion score can be obtained from the R Factor with a conversion formula. For $6.5 < R < 100$, this conversion formula can be inverted:

$$\mathrm{MOS_2R}\left(x\right) = \frac{20}{3}\left(8 - \sqrt{226}\cos\left(h + \frac{\pi}{3}\right)\right) \tag{4}$$

$$\begin{aligned} h = \quad & \tfrac{1}{3}\arctan 2\left(18566 - 6750x,\right. \\ & \left.15\sqrt{-202500x^2 + 1113960x - 903522}\right) \end{aligned}$$

PESQ not verfied ⇨conduct listening tests

- Can PESQ measure playout rescheduling?
- Or non-random packet losses?

- We conducted formal listening-option tests

  Result: Correlation between PESQ and humans is for
  - non random packet losses: R=0.94
  - playout re-scheduling: R=0.87

    (R=0 ⇨ no correlation; R=1 ⇨perfect correlation)

Assessing VoIP: Summary

- Implemented all common playout schedulers
- Verified software in various research projects,
  e.g. voice over WLAN, impact of handover, ad-hoc.
- Open-source software
  http://www.tkn.tu-berlin.de/reseach/qofis

- C. Hoene, S. Wiethölter, and A. Wolisz, "Predicting the Perceptual Service Quality Using a Trace of VoIP Packets", In *Proceedings of QofIS'04*, Barcelona, Spain, September 2004.
- C. Hoene, H. Karl, and A. Wolisz, "A Perceptual Quality Model for Adaptive VoIP Applications", In *Proceedings of International Symposium on Performance Evaluation of Computer and Telecommunication Systems (SPECTS'04)*, San Jose, California, USA, July 2004, Paper won the Best Paper Award of the conference.
- C. Hoene and E. Dulamsuren-Lalla, "Predicting Performance of PESQ in Case of Single Frame Losses", In *Proc. MESAQIN 2004*, Prague, CZ, June 2004.
- S. Möller and C. Hoene, "Information About a New Method For Deriving the Transmissio Rating Factor R From MOS in Closed Form", ITU, May 2002, Temporary Document for the study group 12.

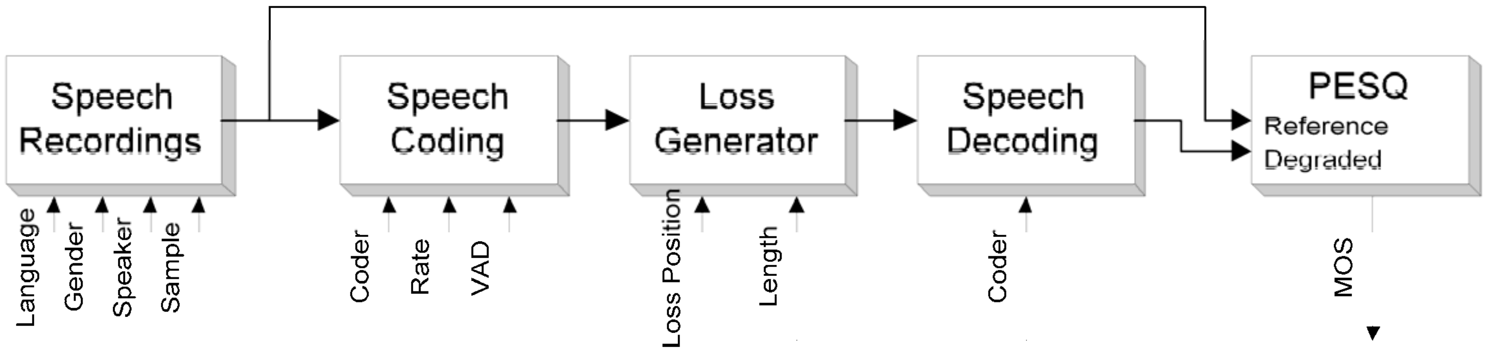
Content

- Motivation: Semantic data-link
- Determining VoIP Quality
  - Perceptual quality of VoIP packet traces
  - Listening-only tests
- On the Importance of a VoIP Packet
  - Audio Demos
  - Predicting the Importance
- Voice over WLAN
  - Experiments: Semantic data-link
  - Simulations: Voice over EDCA
- Conclusions

Introduction

- Losing one Voice-Over-IP packet 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

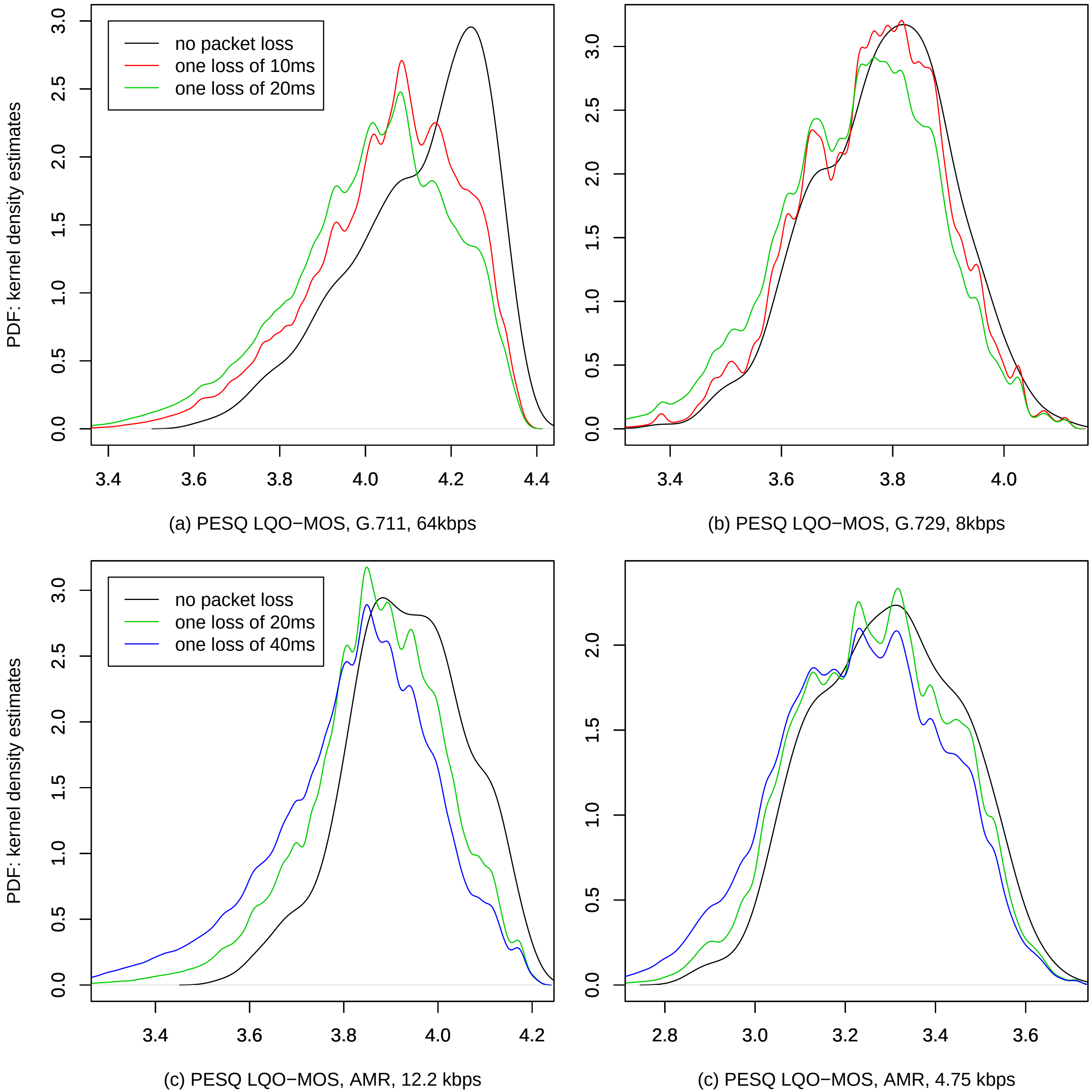
Measurement Setup (1)



- 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)

Measurement Setup (2)

- Loss Generator
  - different positions (800) and loss lengths (1,2,3,4)
- Totally: some millions different tests
- Will take several years
  if humans conduct these speech quality tests.
- Thus, use ITU P.832 PESQ.
- PESQ calculates a Mean Opinion Source


no packet loss
one loss of 10ms
one loss of 20ms
PDF: kernel density estimates
(a) PESQ LQO−MOS, G.711, 64kbps
(b) PESQ LQO−MOS, G.729, 8kbps
no packet loss
one loss of 20ms
one loss of 40ms
PDF: kernel density estimates
(c) PESQ LQO−MOS, AMR, 12.2 kbps
(c) PESQ LQO−MOS, AMR, 4.75 kbps

ion

Definition: 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:

$$\mathrm{Imp}^*\left(s,c,\{l_1,l_2,\ldots\}\right)=\left(\mathrm{MOS}\left(s,c\right)-\mathrm{MOS}\left(s,c,\{l_1,l_2,\ldots\}\right)\right)\cdot \mathrm{t}\left(s\right)$$

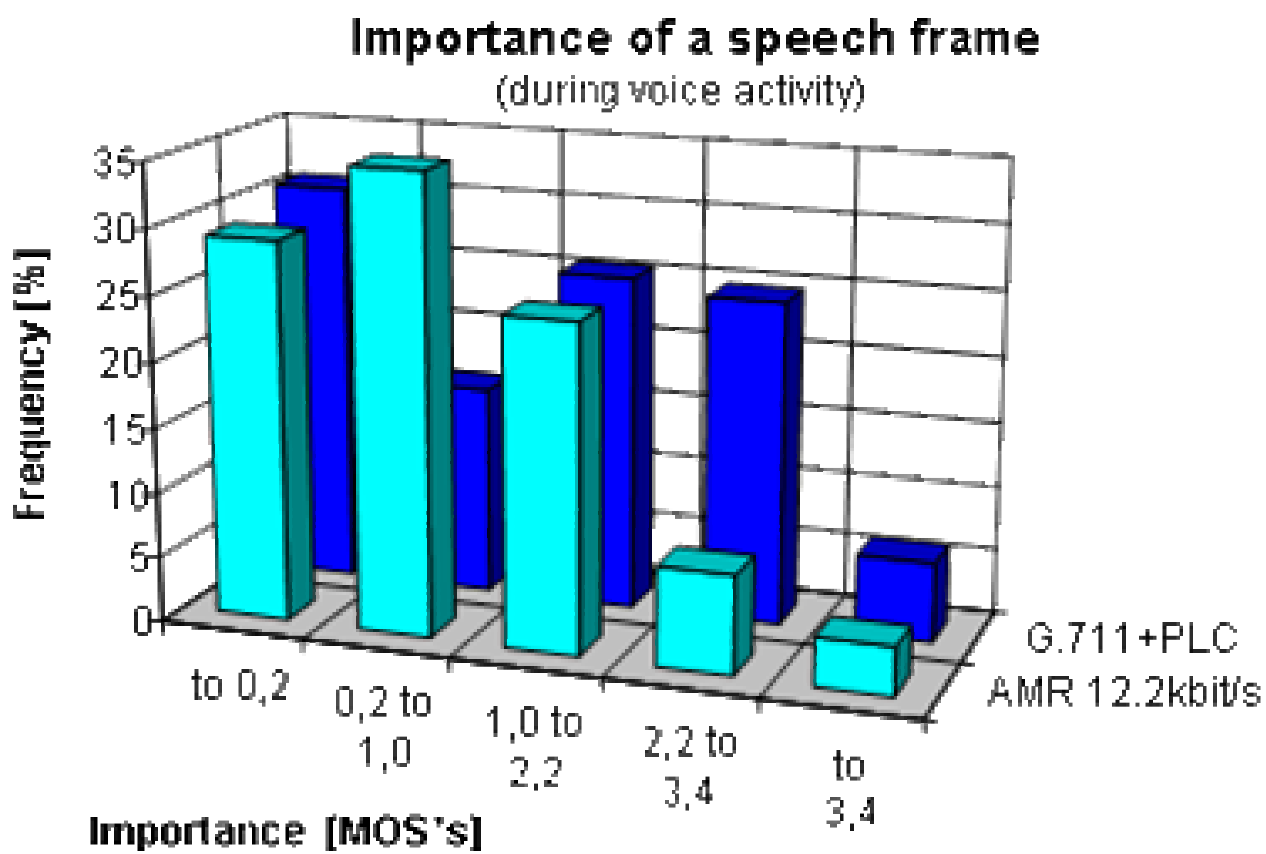
s: sample
t(s): samples length (s)
c: codec implementation
l_x: loss event, one or multiple correlated frame losses

Measuring the Importance of a VoIP packet

- 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

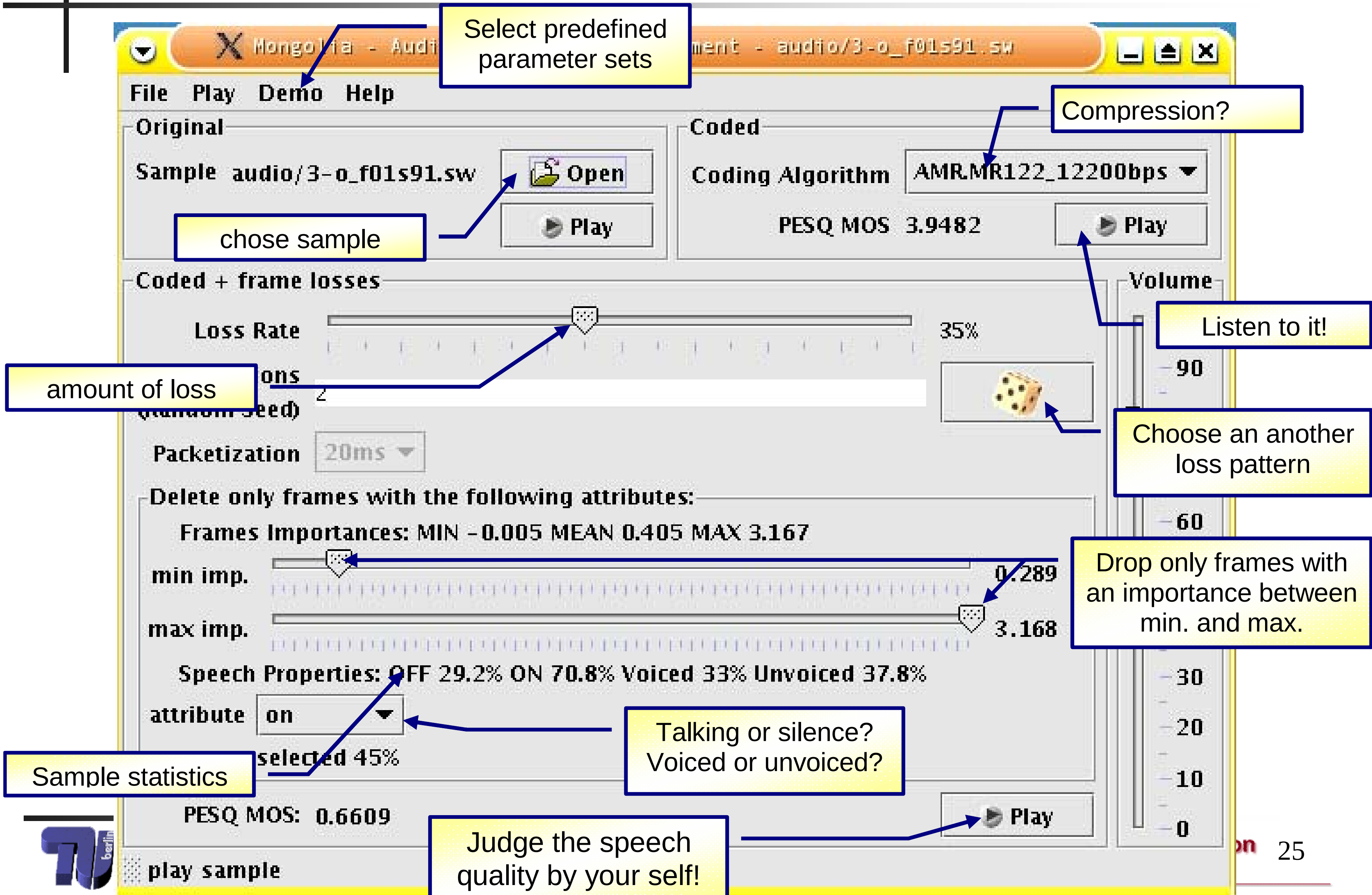
- Can PESQ measure packet importances? Nobody knows…

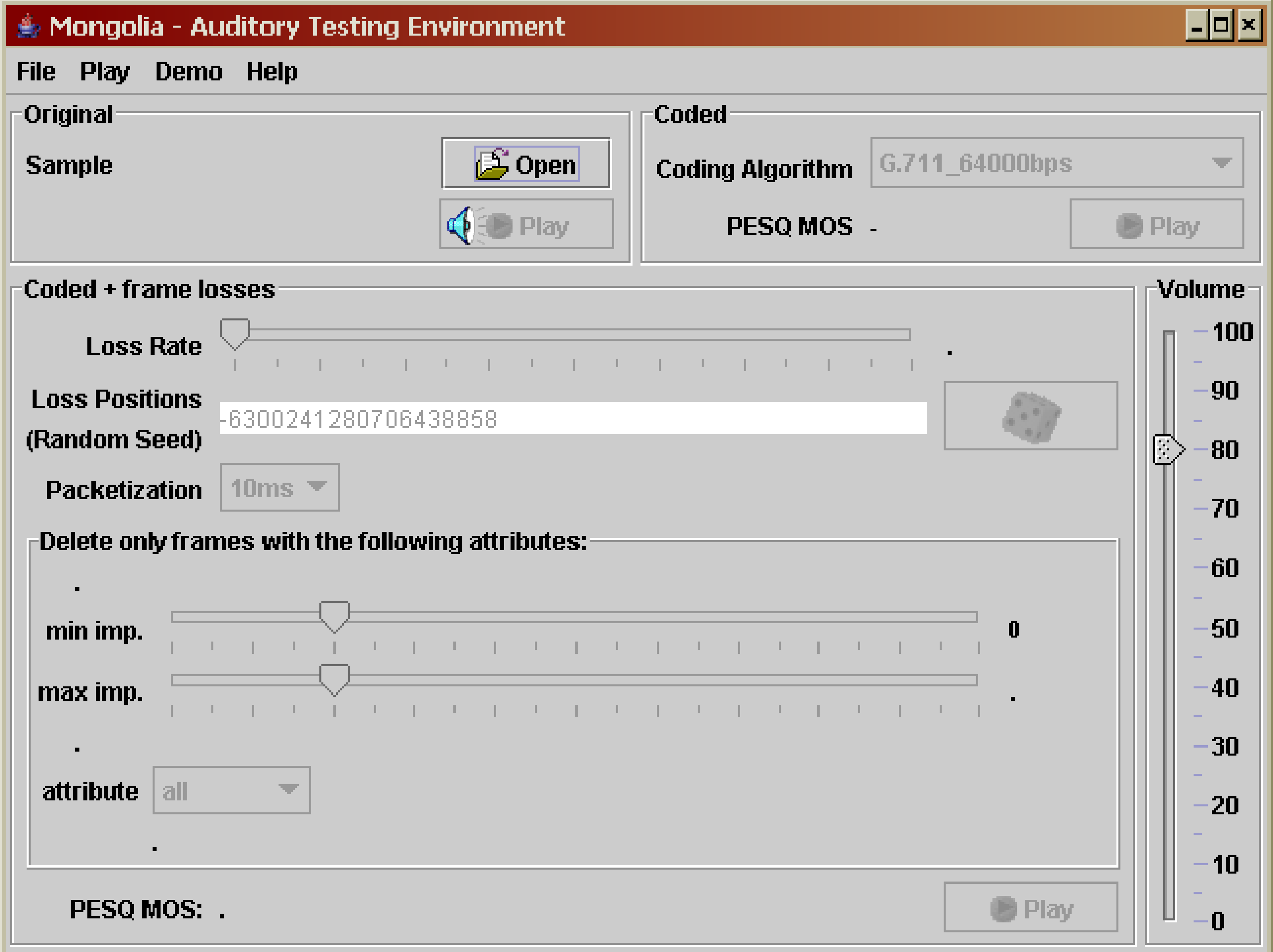
⇨ Thus: Conduct formal listening-only tests!

Problem:
Humans can not hear single packet losses!

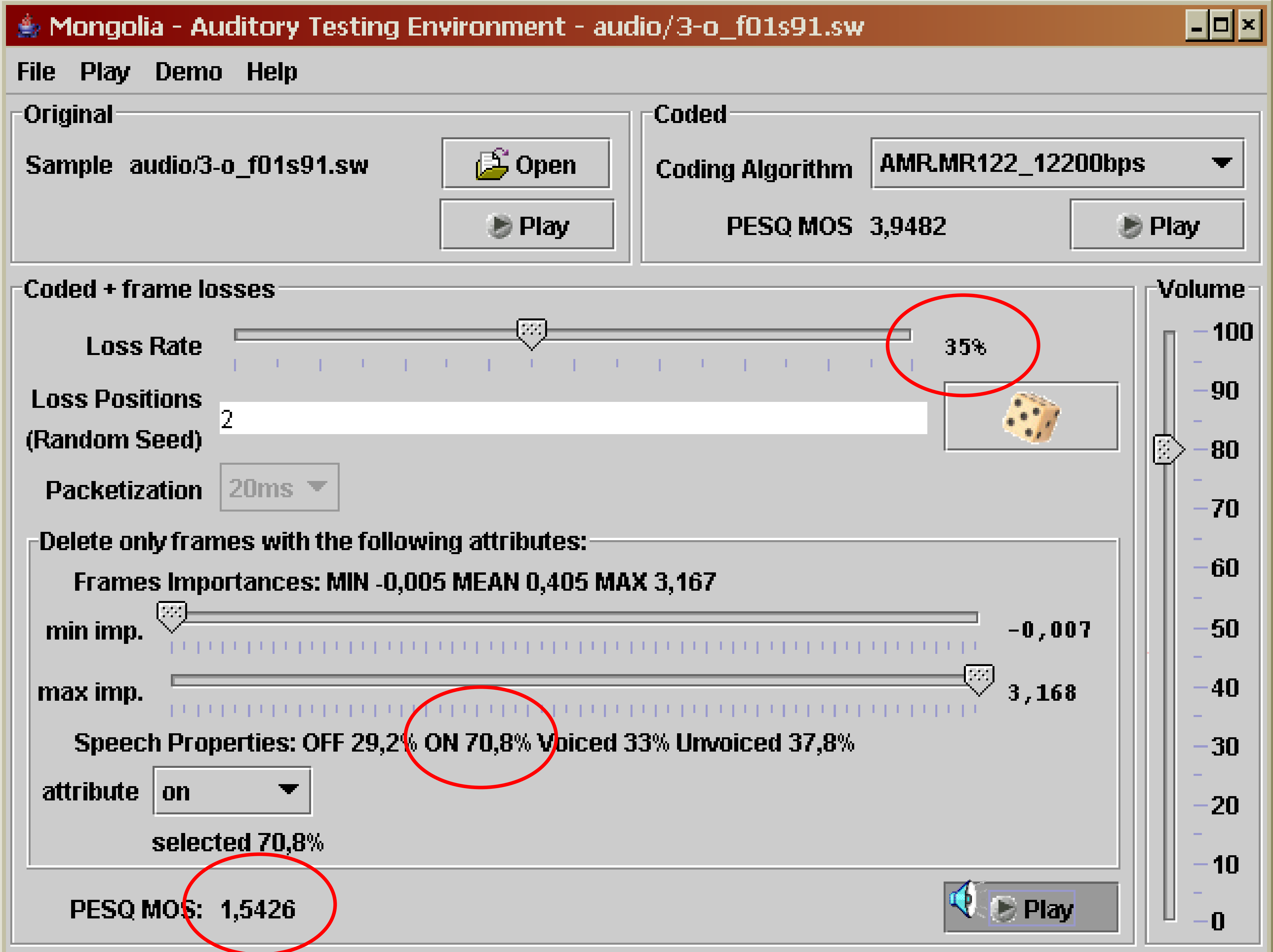
- Human just can hear multiple packet losses.
- Thus, drop multiple similar frames…
- PESQ can identify similar frames.

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

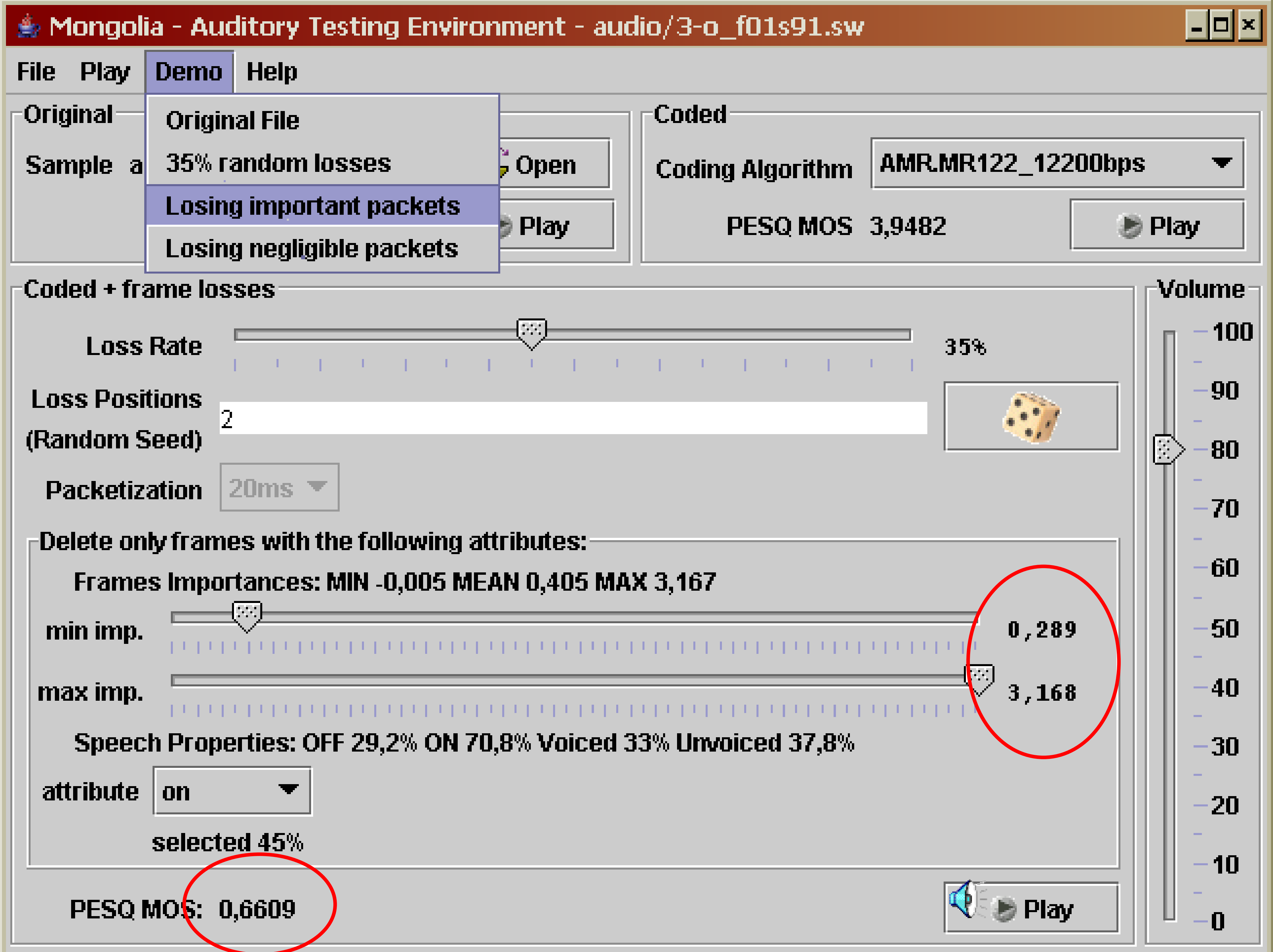


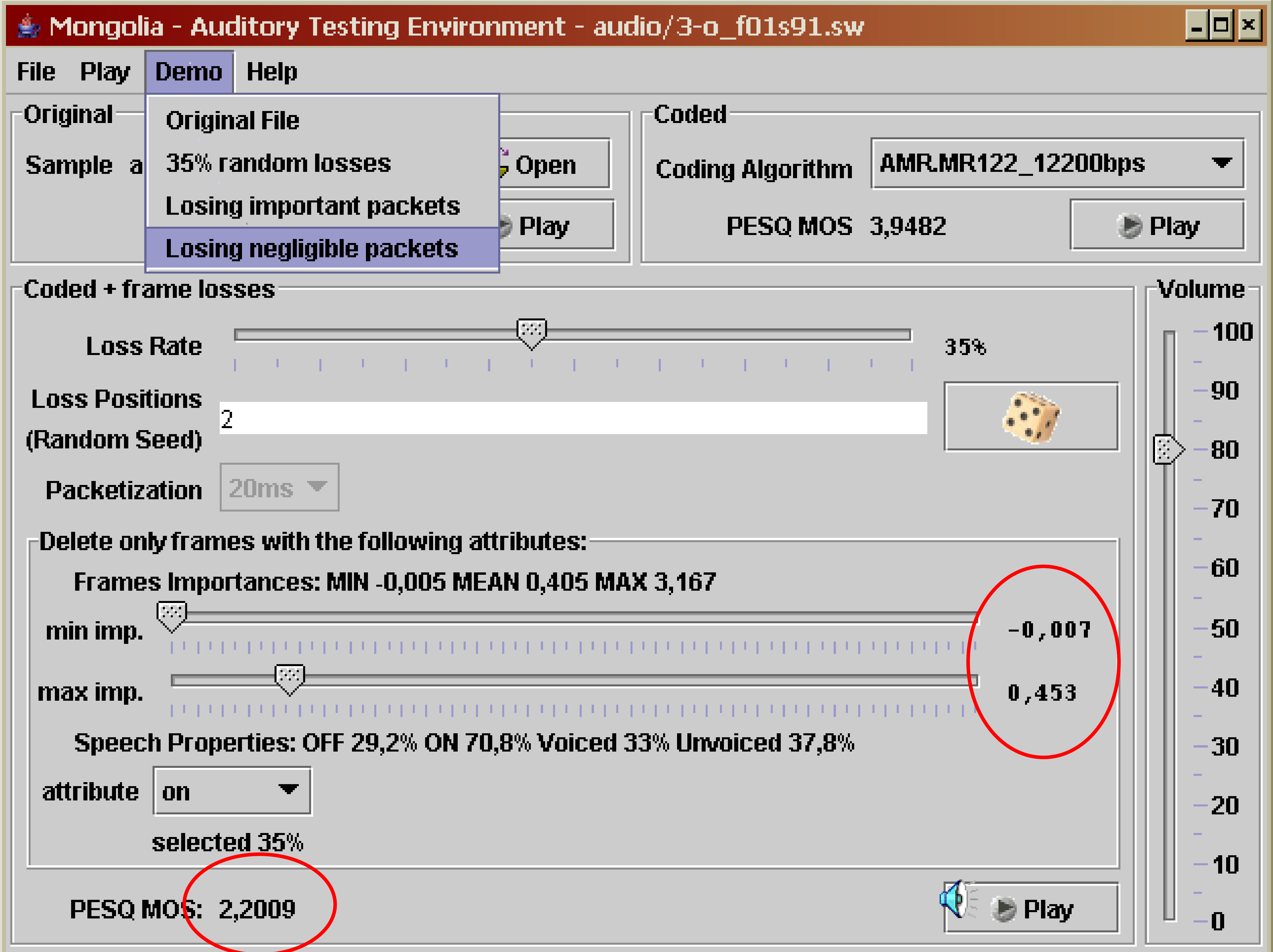


Mongolia - Auditory Testing Environment
File
Play
Demo
Help
Original
Sample
Open
Play
Coded
Coding Algorithm
G.711_64000bps
PESQ MOS -
Play
Coded + frame losses
Loss Rate
Loss Positions
(Random Seed)
-6300241280706438858
Packetization
10ms
Delete only frames with the following attributes:
min imp.
0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

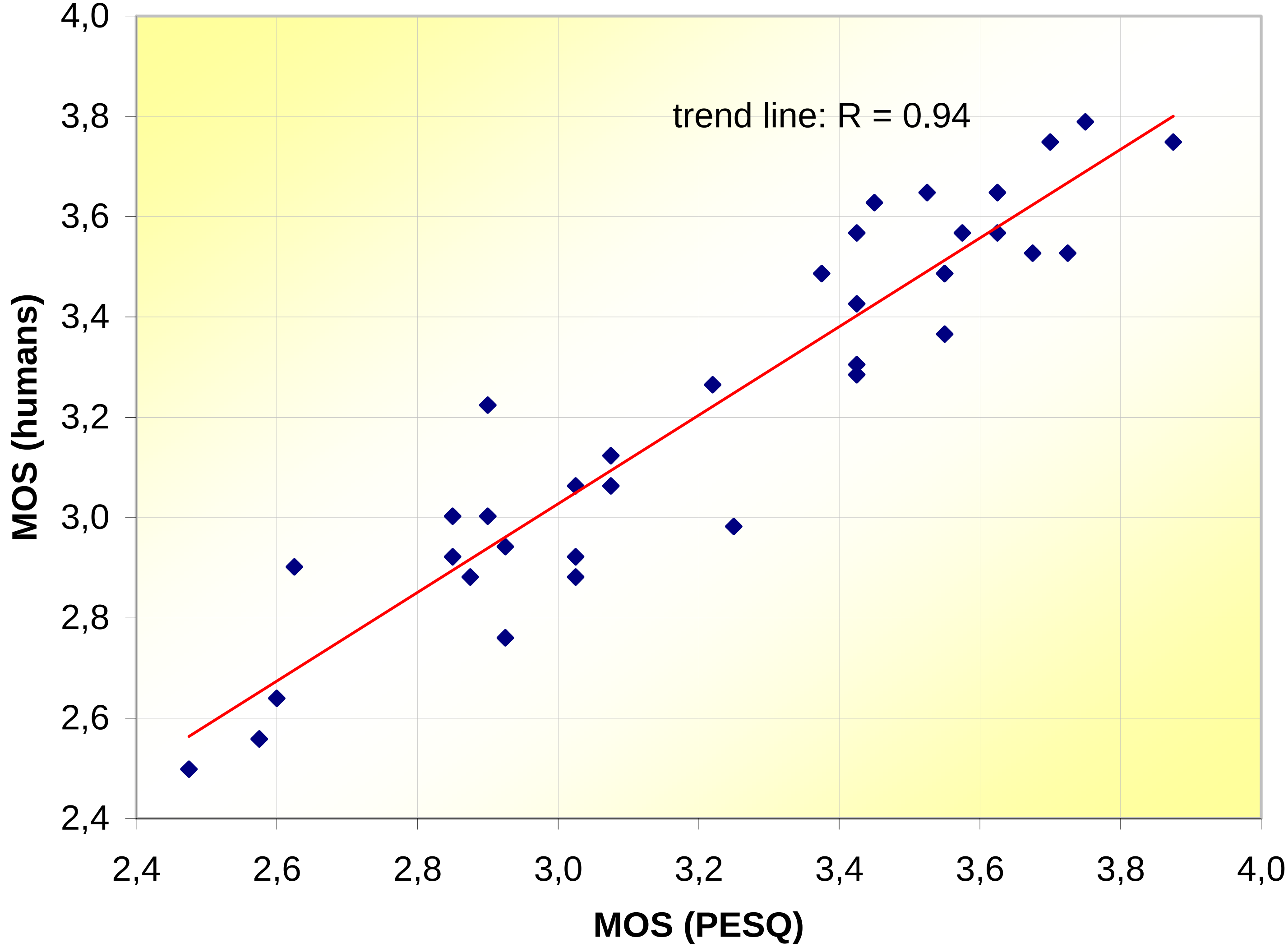


Mongolia - Auditory Testing Environment - audio/3-o_f01s91.sw
File
Play
Demo
Help
Original
Sample audio/3-o_f01s91.sw
Open
Play
Coded
Coding Algorithm
AMR.MR122_12200bps
PESQ MOS 3,9482
Play
Coded + frame losses
Loss Rate
35%
Loss Positions (Random Seed)
2
Packetization
20ms
Delete only frames with the 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

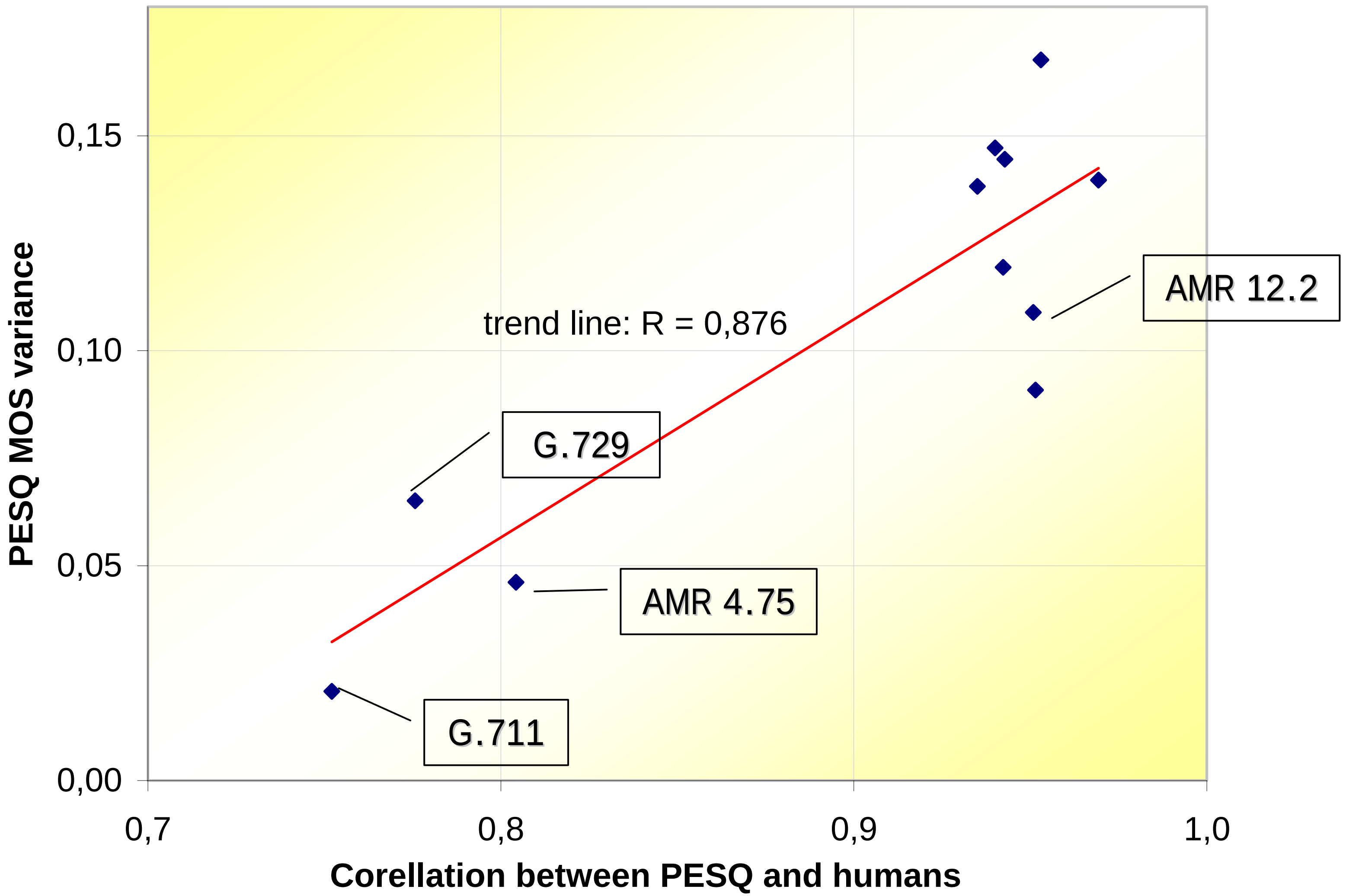

Mongolia - Auditory Testing Environment - audio/3-o_f01s91.sw
File
Play
Demo
Help
Original
Sample a
Original File
35% random losses
Losing important packets
Losing negligible packets
Open
Play
Coded
Coding Algorithm
AMR.MR122_12200bps
PESQ MOS 3,9482
Play
Coded + frame losses
Loss Rate
35%
Loss Positions
(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
Sample a
Original File
35% random losses
Losing important packets
Losing negligible packets
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



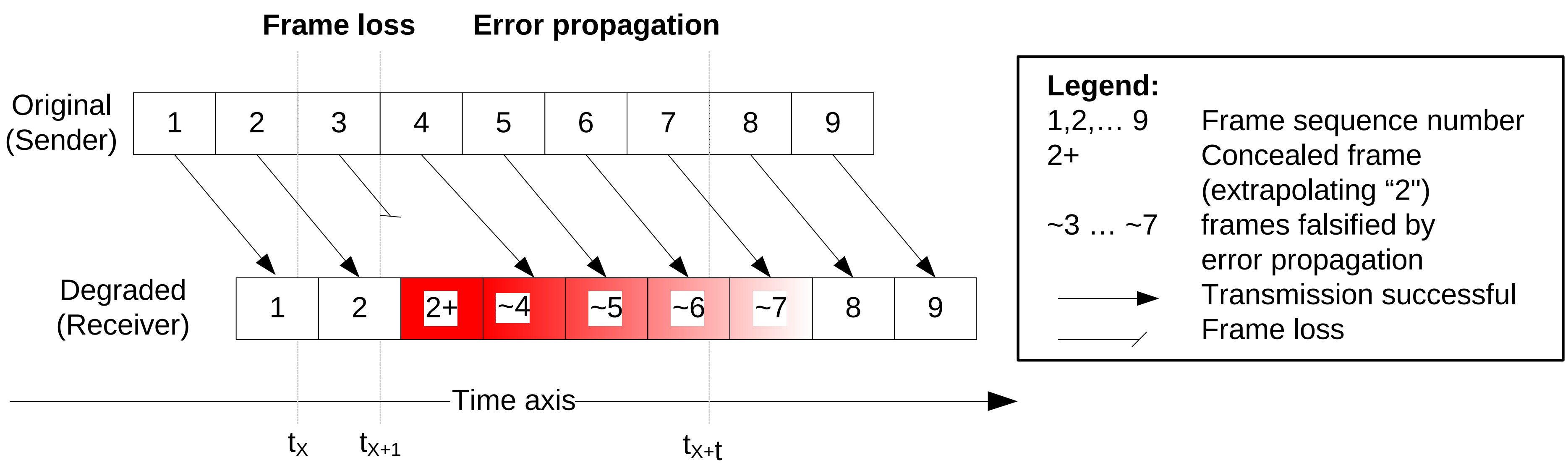
MOS Variance vs. Prediction Performance



Content

- Motivation: Semantic data-link
- Determining VoIP Quality
  - Perceptual quality of VoIP packet traces
  - Listening-only tests
- On the Importance of a VoIP Packet
  - Audio Demos
  - Predicting the Importance
- Voice over WLAN
  - Experiments: Semantic data-link
  - Simulations: Voice over EDCA
- Conclusions

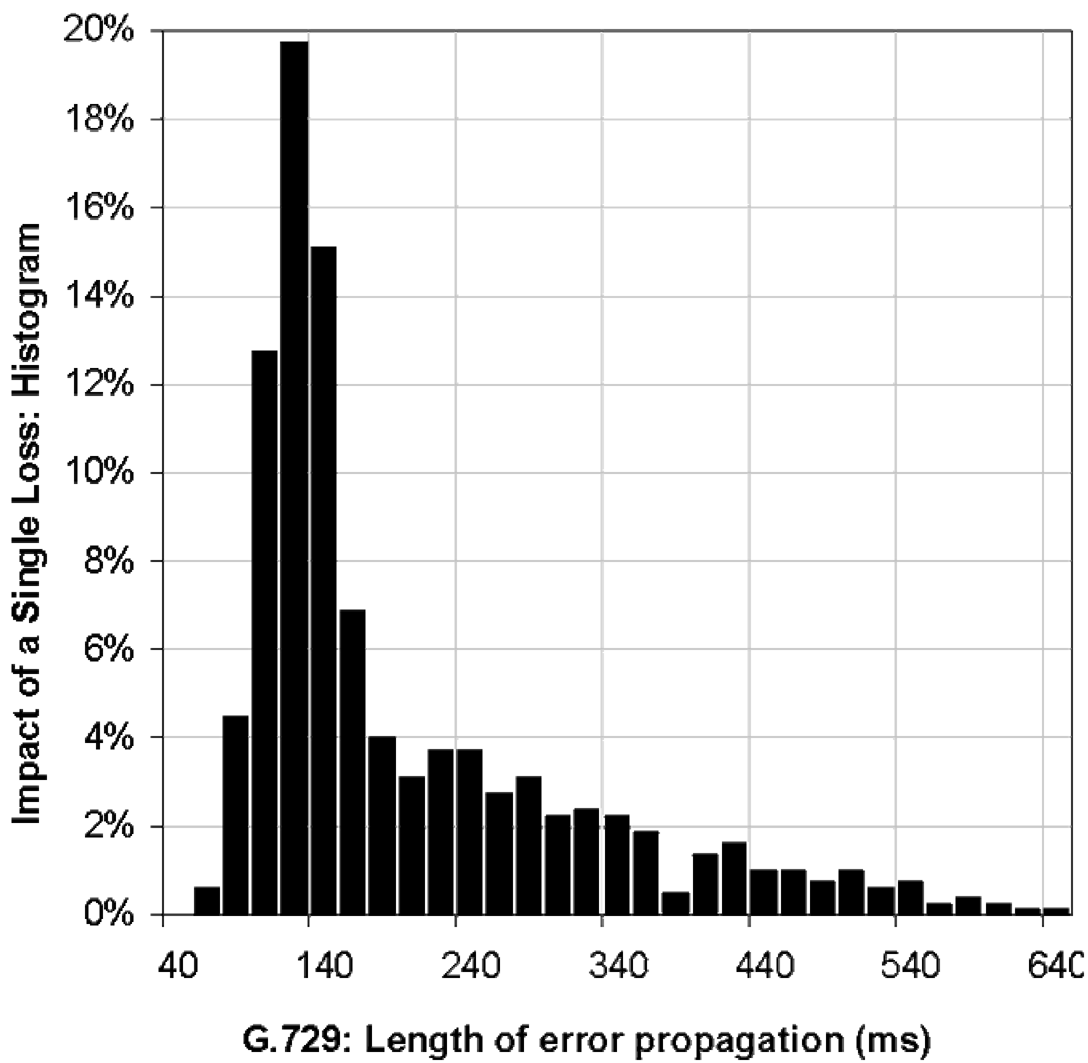
Understanding the Importance



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



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

Impairment	G.711	G.729	AMR12.2	AMR4.75
Left	0.0370 (88%)	0.0431 (86%)	0.1589 (93%)	0.1439 (95%)
Right	0.0048 (12%)	0.0068 (14%)	0.0117 (7%)	0.0070 (5%)
Both	0.0419	0.0499	0.1706	0.1509

- Importance of a Speech Frame depends on
  - Imperfect concealment (9/10)
  - Error propagation (1/10)
- Good news, because
  - Imperfect concealment can be measured
  - Error propgation cannot be measured

Predicting the Importance at Real-Time

- Algorithms by Sanneck, De Martin, Hoene
- Assessment, analysis, more details to come
- (Not yet published)

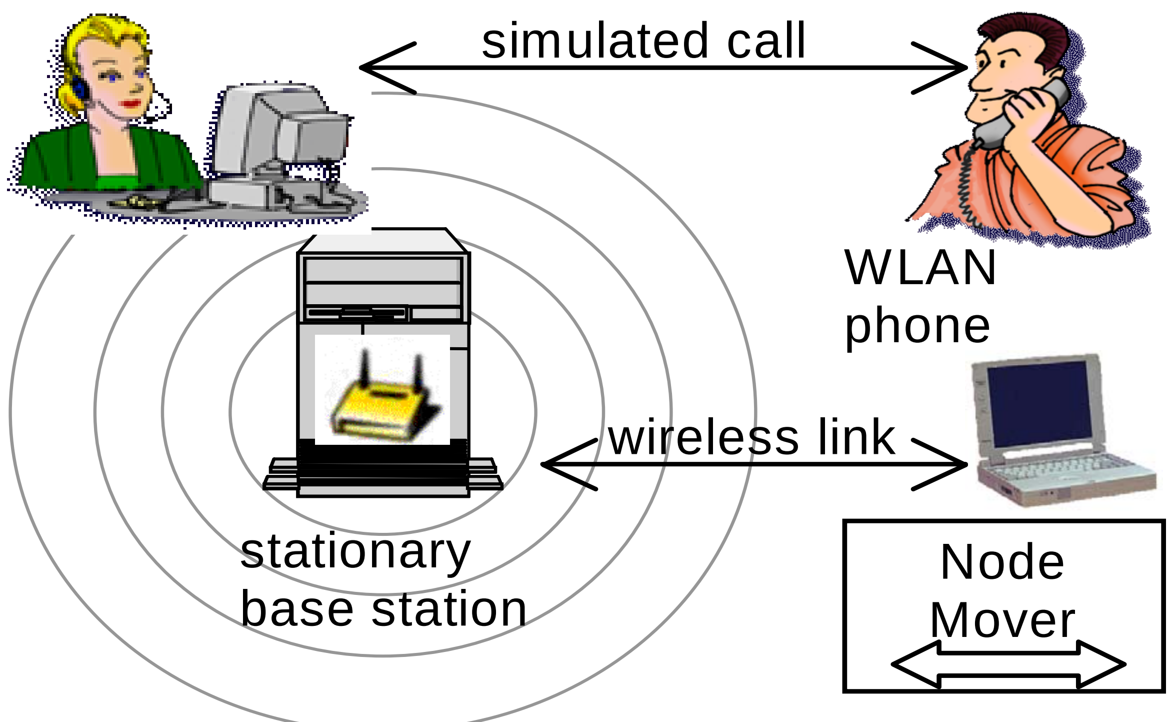
Content

- Motivation: Semantic data-link
- Determining VoIP Quality
  - Perceptual quality of VoIP packet traces
  - Listening-only tests
- On the Importance of a VoIP Packet
  - Audio Demos
  - Predicting the Importance
- Voice over WLAN
  - Intermezzo
  - Experiments: Semantic data-link
  - Simulations: Voice over EDCA
- Conclusions

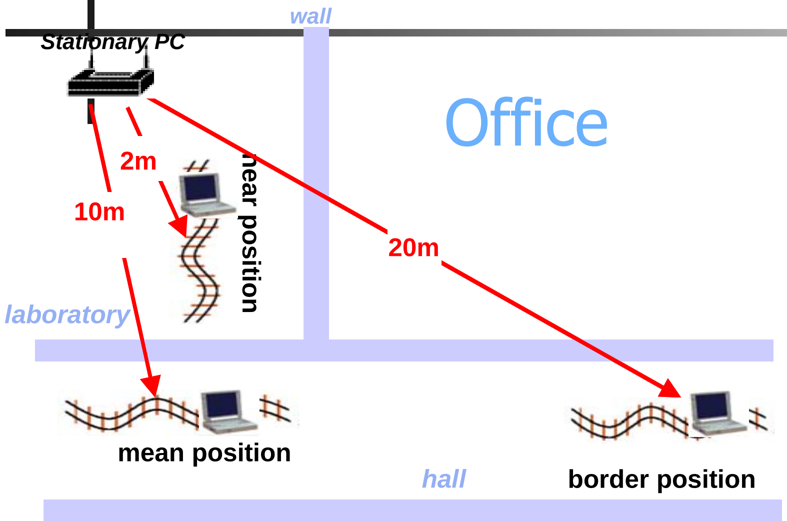
Impact of Movement on Link Quality?

Measured loss rate and TX delay

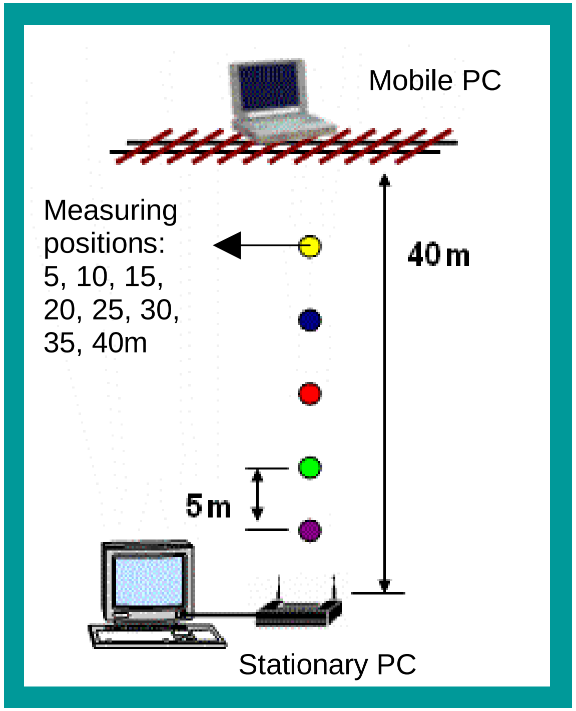
- A bidirectional phone call, consisting solely of RTP packets
- Communication between an access point and mobile node
- Experimental enforced controlled motion (Node Mover)

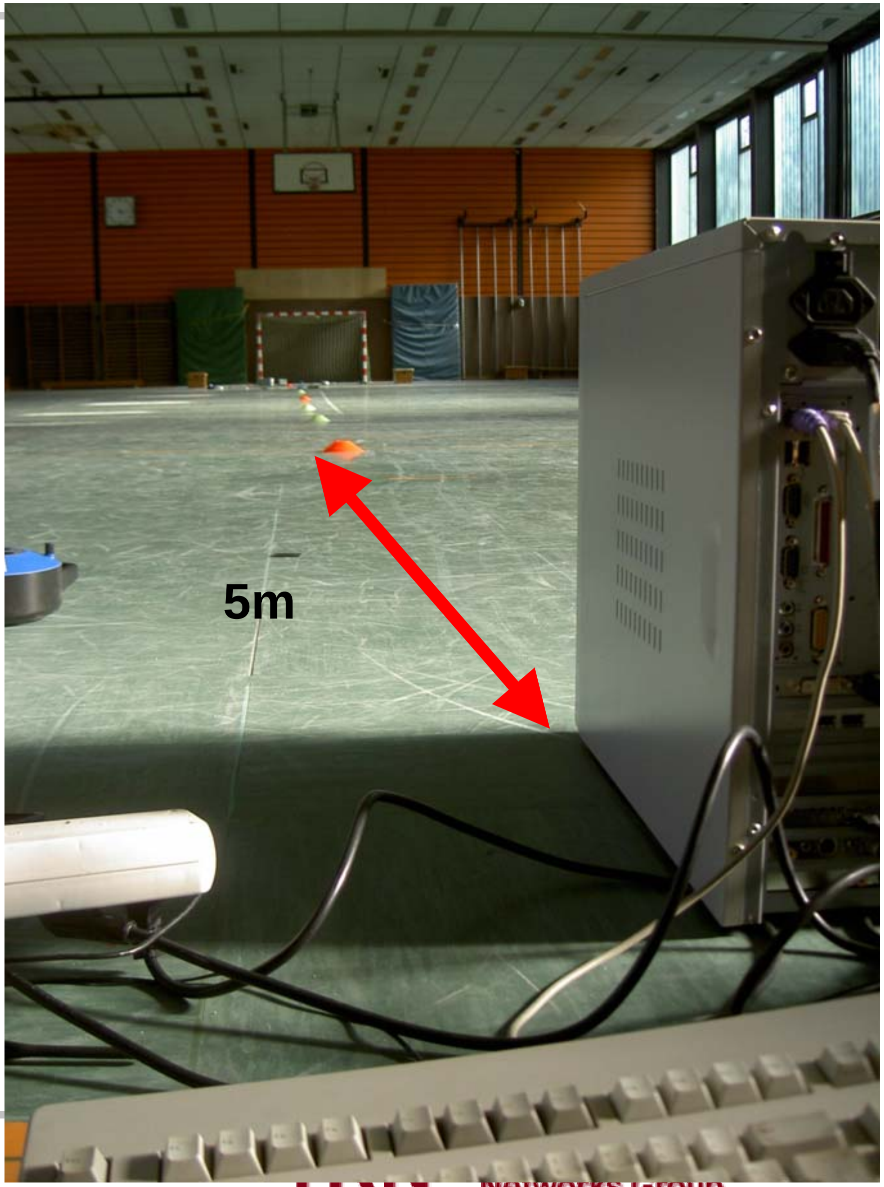


Measurement Environments (Office and Gym)

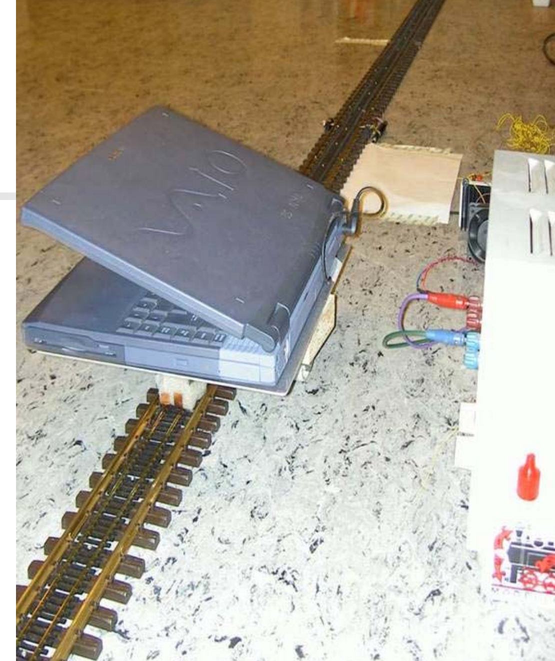


Gym:

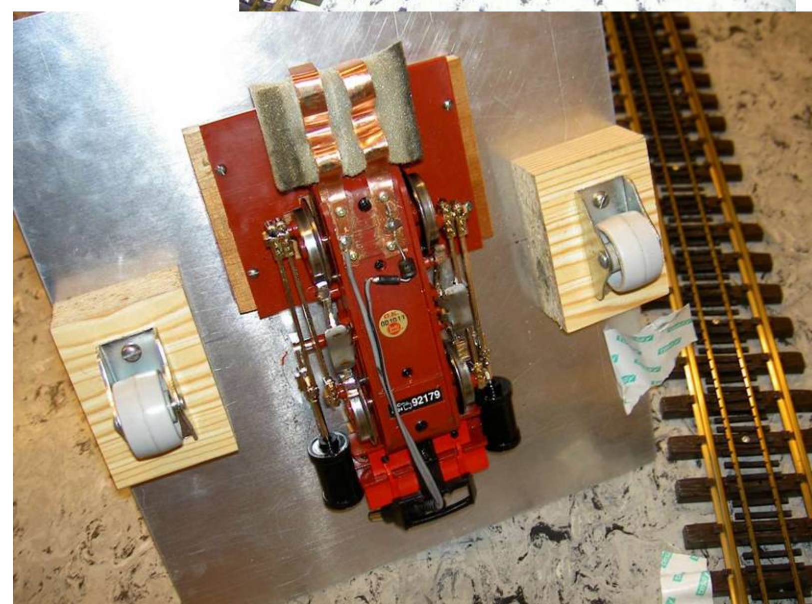




Node Mover (1)

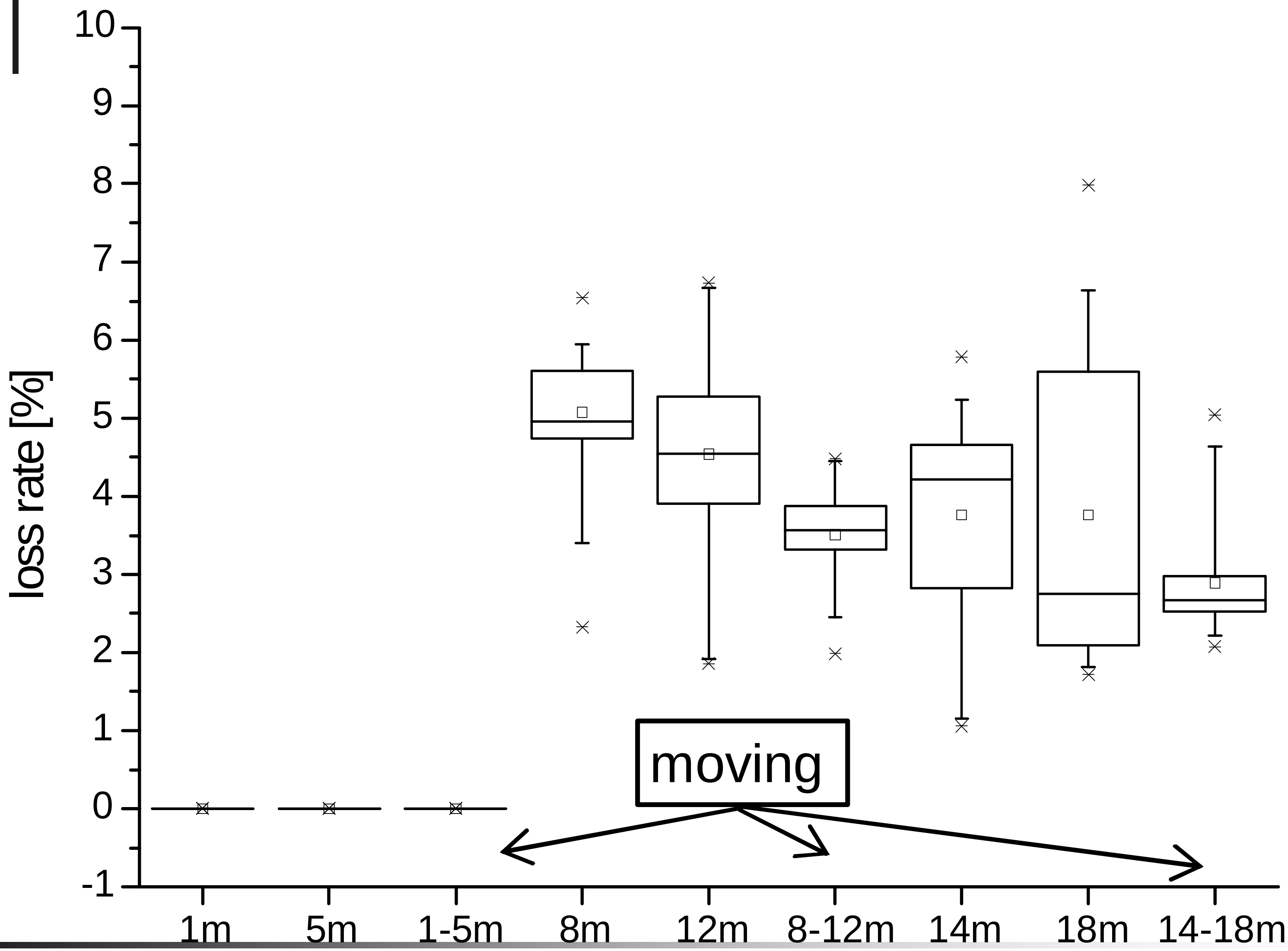


- How to enforce controlled motion? reproducable, over long time spans?
- Human based experiments not feasible…
- Solution: A prepared toy train
- How to supply notebook with power?
  1. Separate Cable
  2. Supplementary, power carrying rails

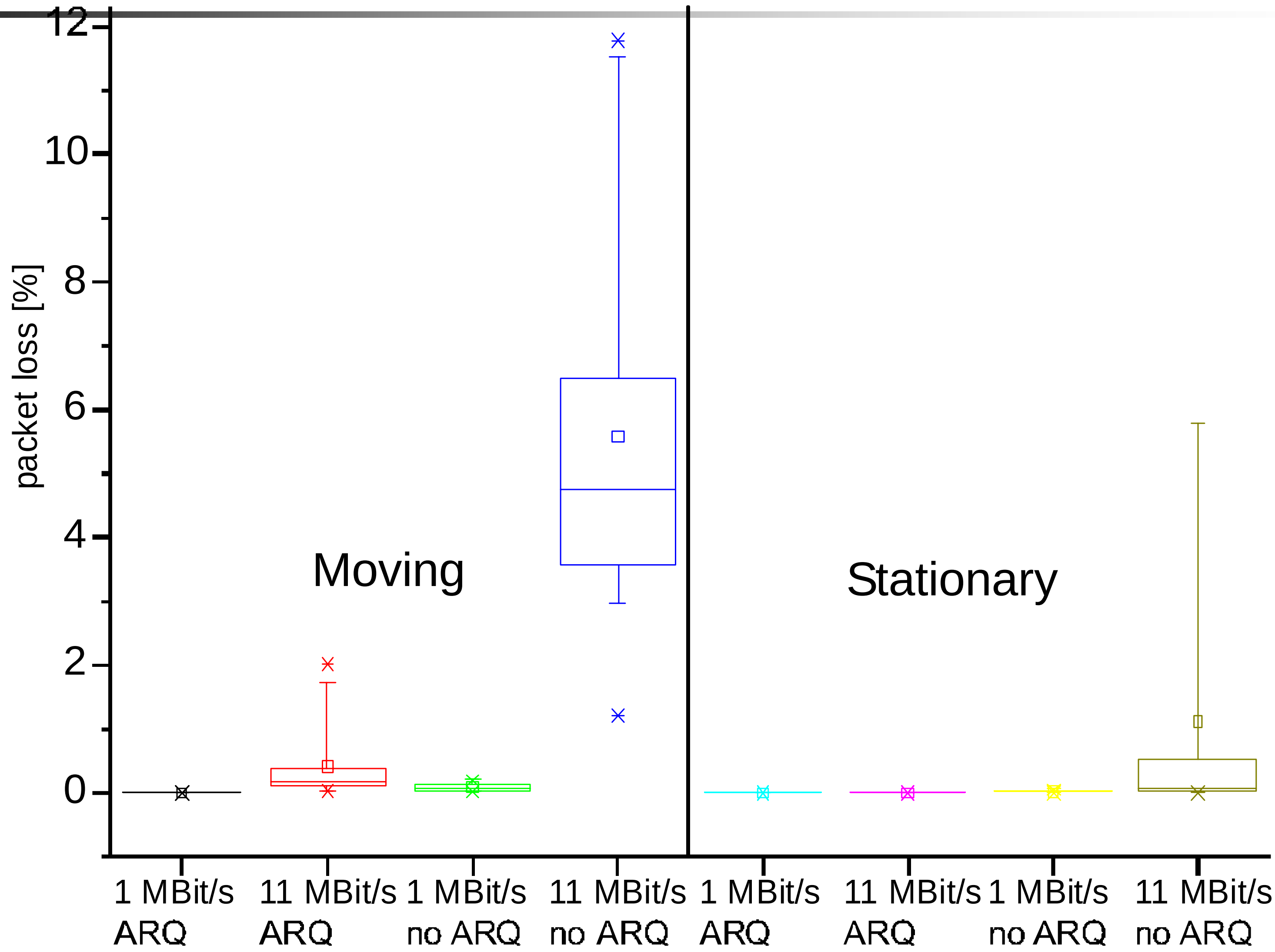


Node Mover (2)

1: Office Environment using vanilla RM conf.



3: Overall loss rate (mean from all experiments)



Summary

- We presented experimental set-up and methodology measuring IEEE 802.11b wireless link
- Develop software Snuffle, which traces protocol messages as well as states on multiple layers.
- Measured packet loss and transmission delay, influenced by modulation type, ARQ, environment, slow motion, and others factors.
- Slow motion usually decreases link quality
- In a Rayleight fading environment using automatic rate selection loss rate were lower! Mere chance?

Introduction: QoS in WLAN

- **How to provide time-bounded services like VoIP / audio / video streaming over WLANs?**

  ➔ hard requirements: low delay, low jitter and constant throughput

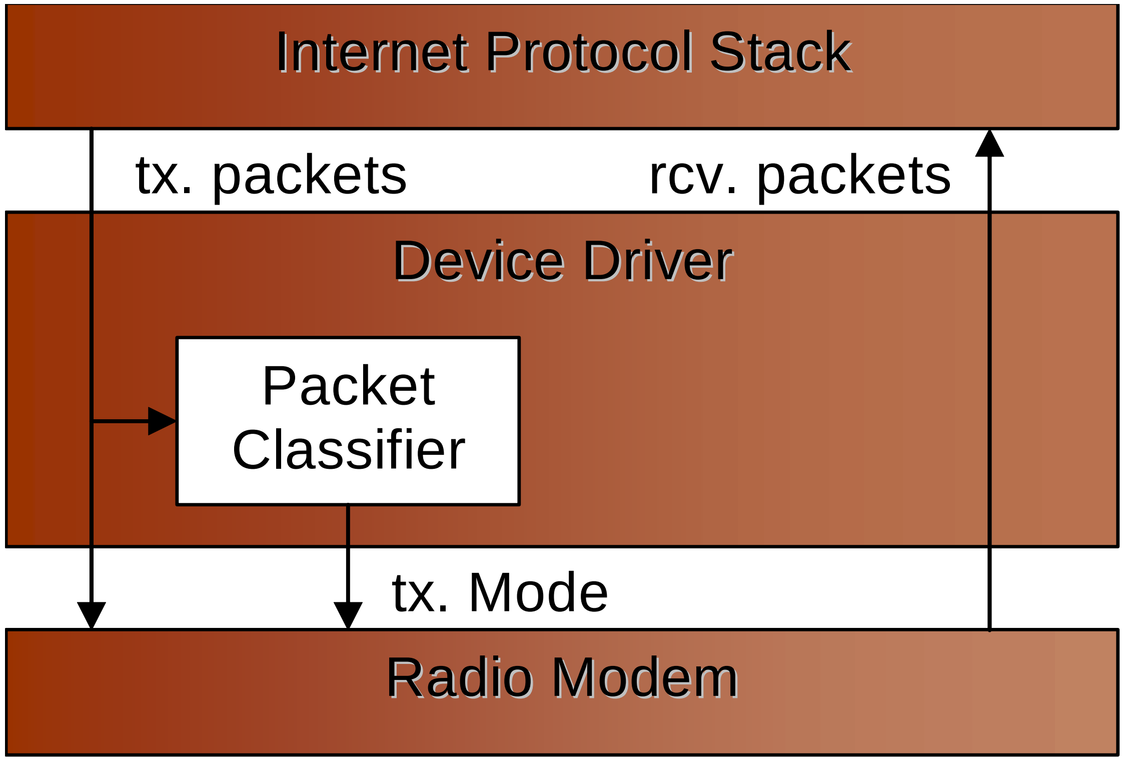
- **Medium access in legacy 802.11:**
  - point coordination function (PCF)
  - distributed coordination function (DCF)

- **Medium access in 802.11e:**
  - hybrid coordination function (HCA)
  - enhanced DCF (EDCA)
  - packet frame grouping (CFB, not include in std. anymore)

Content

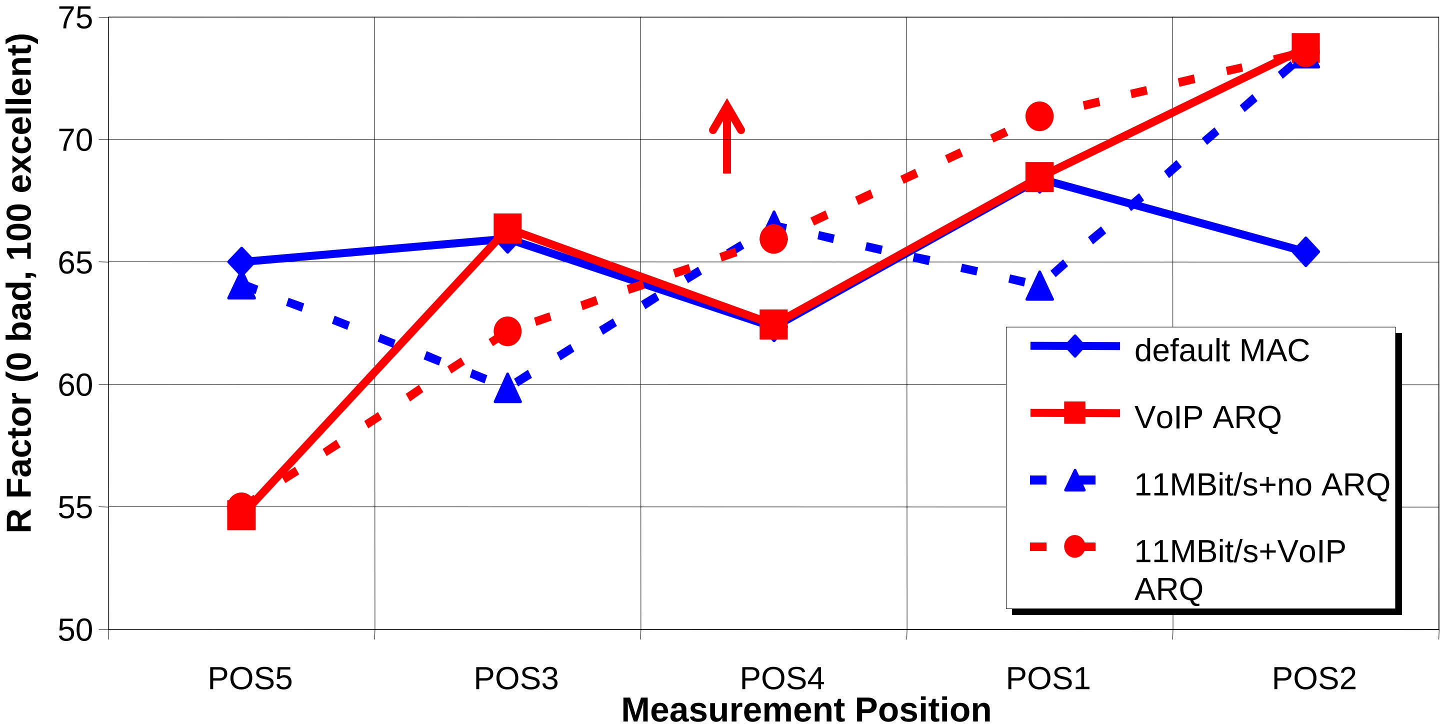
- Motivation: Semantic data-link
- Determining VoIP Quality
  - Perceptual quality of VoIP packet traces
  - Listening-only tests
- On the Importance of a VoIP Packet
  - Audio Demos
  - Predicting the Importance
- Voice over WLAN
  - Experiments: Semantic data-link
  - Simulations: Voice over EDCA
- Conclusions

Adaptive Data Link: Implementation



- Feasibility study and prototype
- Voice sounds are protected by a higher number of retransmissions
- Changed device driver of 802.11b WLAN card
- URL http://www.tkn.tu-berlin.de/research/easysnuffle/

Semantic Data-Link Protocol



- Often improvement of speech quality, even when using a simple algorithm
- How large is the gain that can be reached with a sophisticated algorithm?

Content

- Motivation: Semantic data-link
- Determining VoIP Quality
  - Perceptual quality of VoIP packet traces
  - Listening-only tests
- On the Importance of a VoIP Packet
  - Audio Demos
  - Predicting the Importance
- Voice over WLAN
  - Experiments: Semantic data-link
  - Simulations: Voice over EDCA
- Conclusions

Simulating IEEE 802.11e

- Implementation of a simulation model supporting EDCF and CFB
  for the ns-2.26 network simulator
- Subset: Correctness is tested
- Mangold simulator is not free available.
- Repeated the research of Mangold et al. (RWTH-Aachen) ➔ no major difference
- Open-source software:
  500 downloads since December 2003

  http://www.tkn.tu-berlin.de/research/802.11e_ns2/

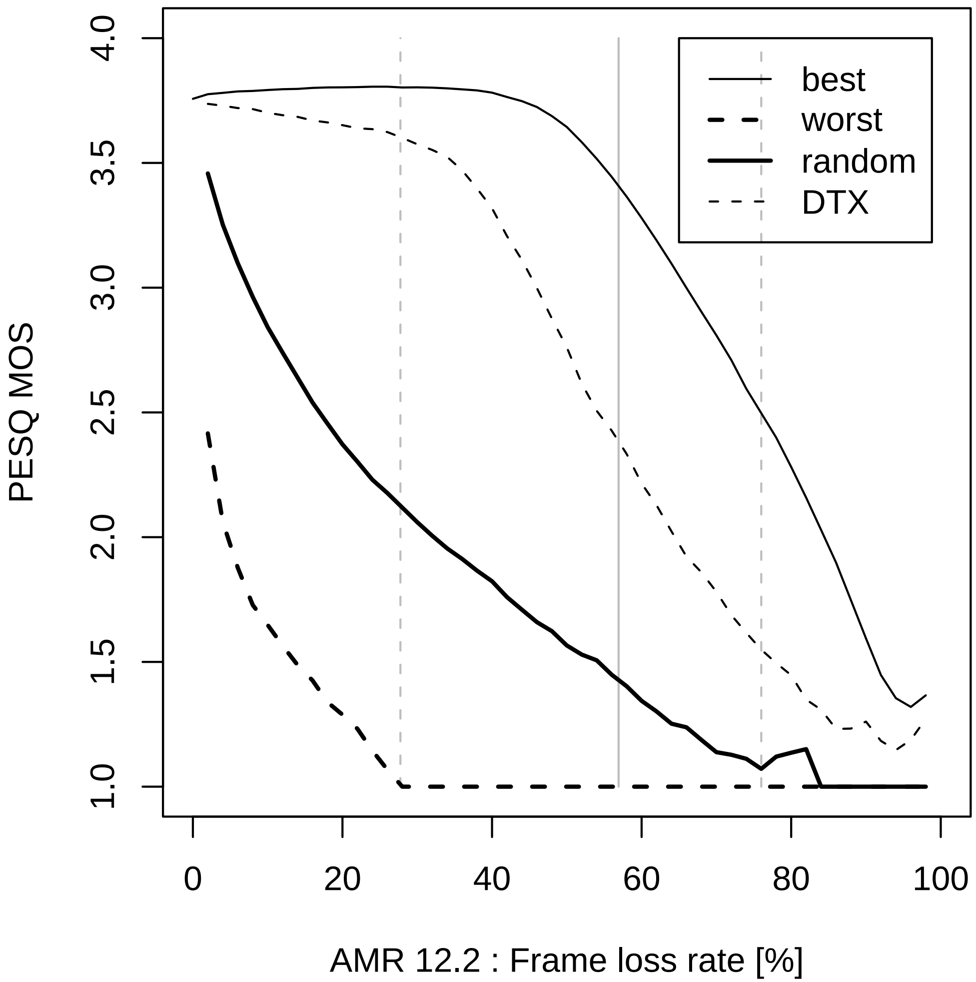
Voice over IEEE 802.11e EDCA

- Open source, modular and verified IEEE 802.11e simulation model for ns-2.26
- VoIP Results
  - Strong improvements with EDCA, EDCA+CFB
  - Low costs for EDCA compared to DCF
  - Max. number of VoIP calls with DCF: 10; EDCA: 10, EDCA + CFB: 11

➔Simple EDCA is adequate for voice calls and

➔HCA only for a high number of calls and for the downlink.

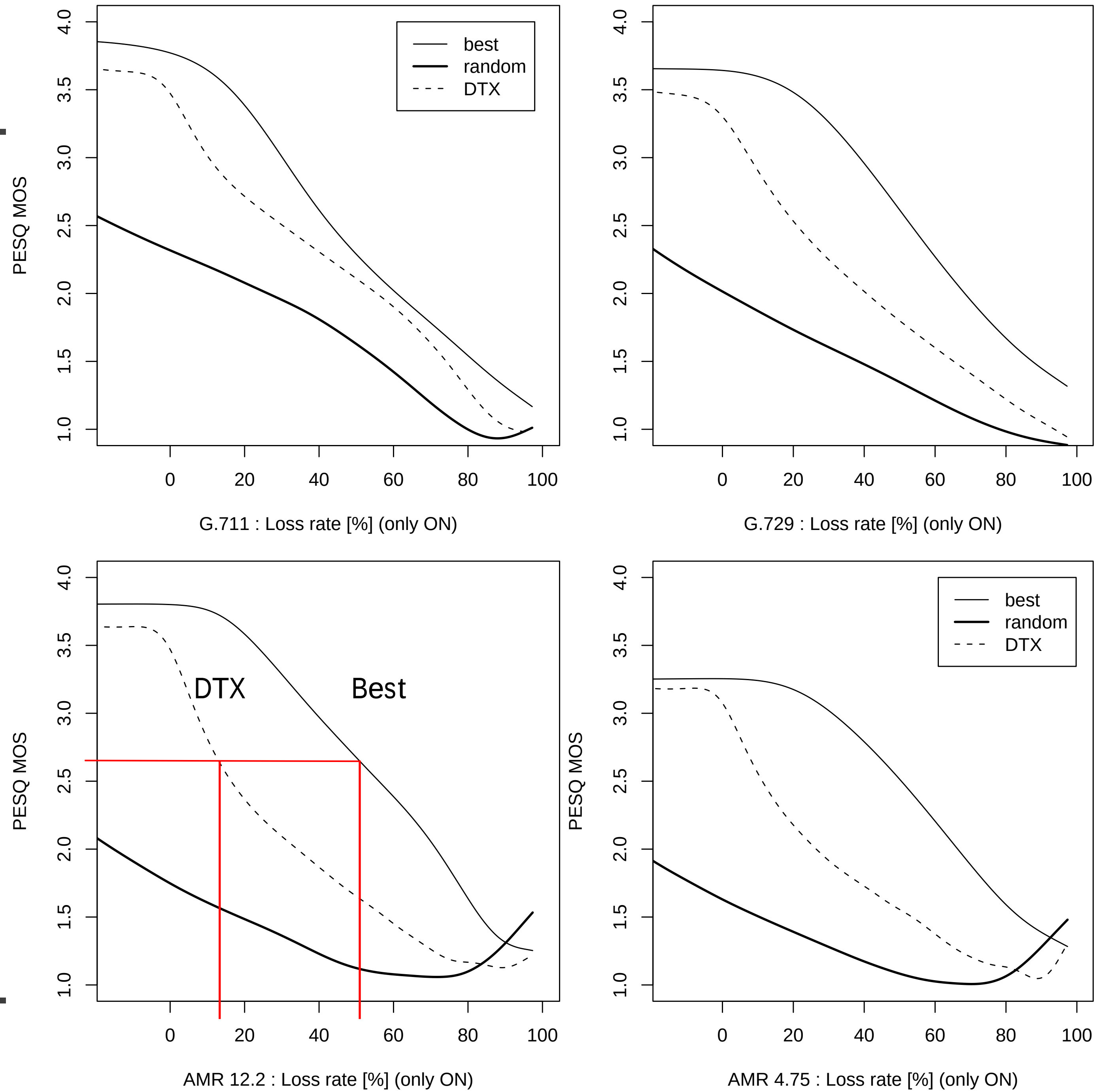
Speech Frames – Dropping Strategy



Drop packets in cases of

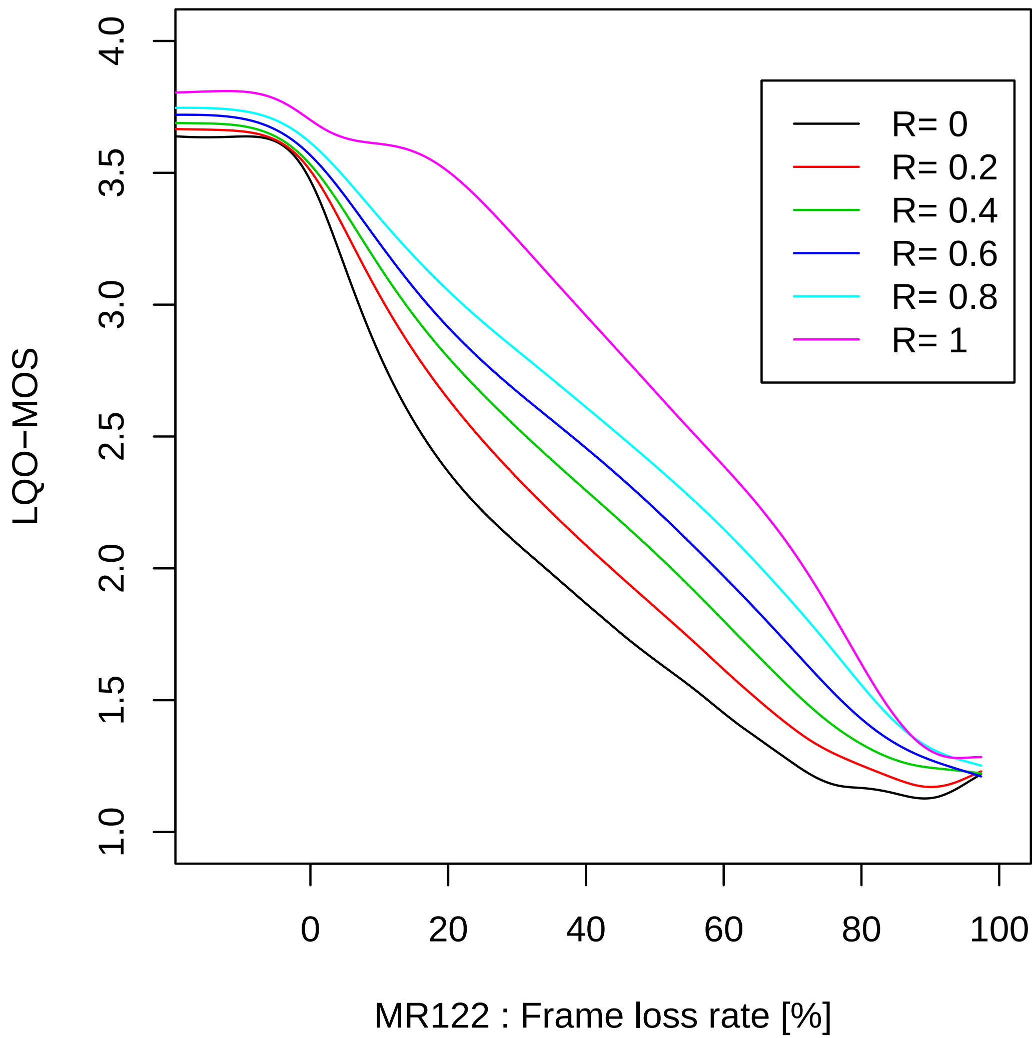
- Congestion
- Wireless fading
- Saving Energy

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



best
random
DTX
PESQ MOS
4.0
3.5
3.0
2.5
2.0
1.5
1.0
0
20
40
60
80
100
G.711 : Loss rate [%] (only ON)
G.729 : Loss rate [%] (only ON)
DTX
Best
AMR 12.2 : Loss rate [%] (only ON)
AMR 4.75 : Loss rate [%] (only ON)

Impact of Prediction Error
(Importance is Falsified by Random Noise)



Conclusion

- Enhancing speech quality on
  VoIP over WLAN
  with semantic data-link protocol
  supporting packet importances

- Packet importance for GSM an UMTS to reduce power consumption?

- Thanks to our hard-working students

- Questions?