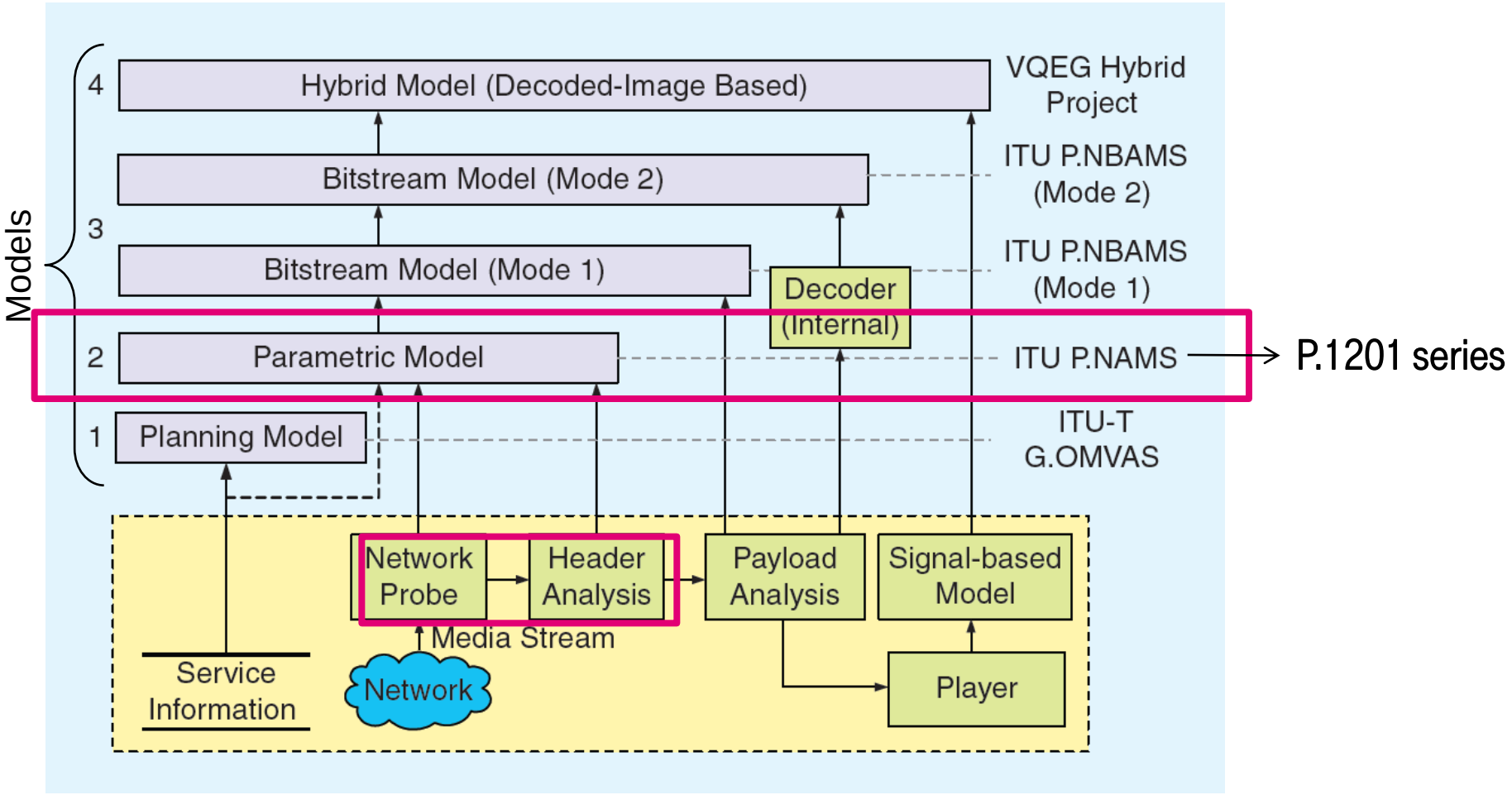
ITU-T Rec. P.1201: standardized parametric packet-based model for audiovisual quality assessment in IPTV and Progressive Download services

M.-N. Garcia[1, 2], P. List[2], S. Argyropoulos[1,2], B. Feiten[2], A. Raake[1,2]
[1]Technische Universität Berlin, [2]Telekom Innovation Laboratories

P.1201 series within the multi-layer model framework



(Raake et al., IEEE SPM 2011)

P.1201 series overview (1/2)

- Covered services
  - P.1201.1:
    - QCIF, QVGA, HVGA, mostly for mobile TV and streaming
    - Protocol: RTP/UDP/IP
  - P.1201.2:
    - SD, HD720, and HD1080 TV, mostly for IPTV
    - Protocol : (MPEG2-TS)/(RTP)/UDP/IP
  - P.1201 Appendix III (P.1201-PD, formerly PNAMS-PD):
    - Use of P.1201 for non-adaptive, progressive download type media streaming; trained on HVGA and HD1080, can be used with P.1201above resolutions.
    - Protocol: TCP/IP

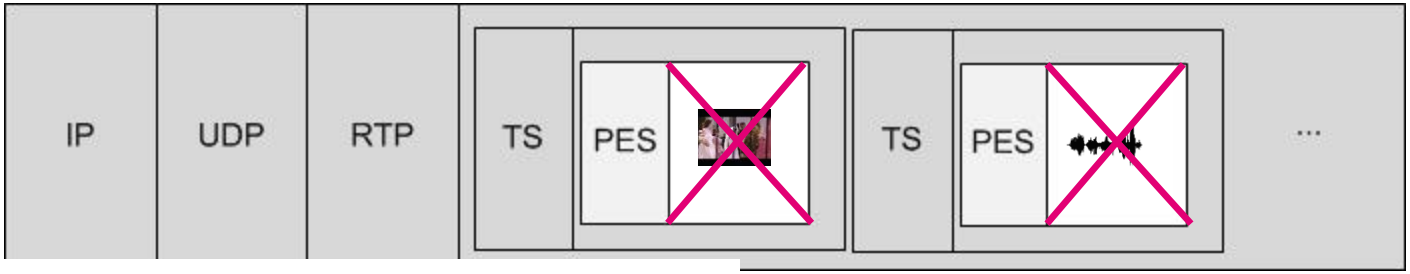
- P.1201 series proponents (alphabetic order)
  - P.1201.1 and P.1201.2: Deutsche Telekom, Ericsson, Huawei, Netscout, NTT, Technicolor, Yonsei university
  - P.1201-PD: Avvasi, Deutsche Telekom, Ericsson, Huawei, Netscout

P.1201 series overview (2/2)

- Packet-header-based models
- Provide estimates of audio, video, and audiovisual quality (MOS) scores

	Video coding	Audio coding	Initial loading	Freezing with skipping	Freezing without skipping	Slicing	Audio packet loss
P.1201.1	MPEG4 VSP H.264 baseline profile, various GOP structures	AMR-NB, AMR-WB+, AAC-LC, HE-AAC	no	yes	yes	1 slice/frame, 1 slice/packet	Yes
P.1201.2	H.264 main & high profiles, various GOP structures	AAC-LC, HE-AAC, MPEG1-LII, AC3	no	no	yes	1 slice/frame, 1 slice / Macroblock row	Yes
P.1201-PD	see above	see above	yes	yes	no	no	no

P.1201.1 and P.1201.2 inputs & outputs



Encrypted PCAP (10-16s sequence)

Model

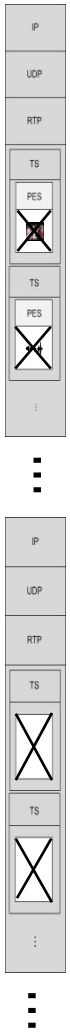
Audio, video, audiovisual MOS

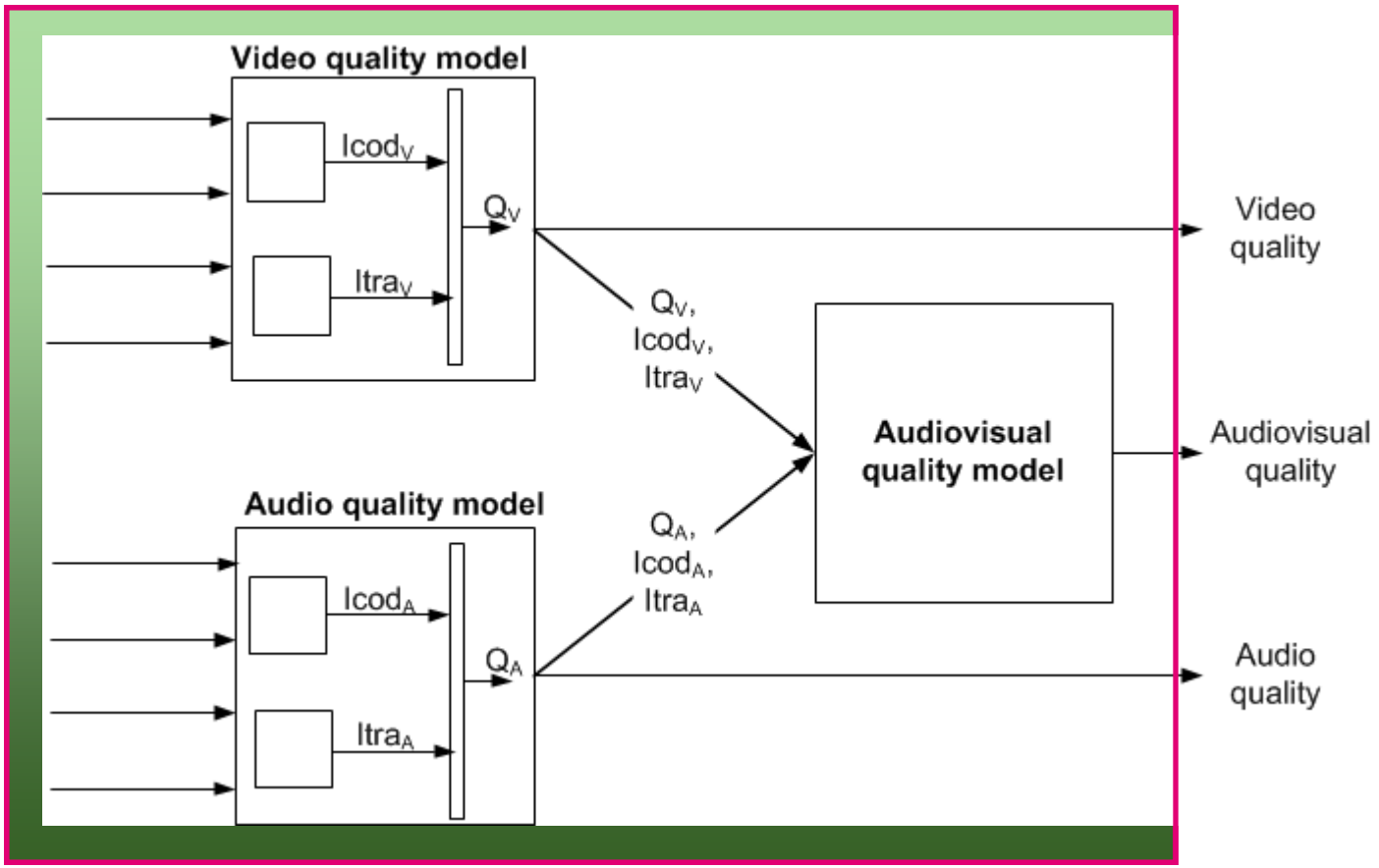
Diagnostic information (e.g. $Icod_A$, $Icod_V$, $Itra_A$, $Itra_V$, see next slide)

Static side information (e.g. audio/video codec, packet-loss-concealment)

Stalling information (start and duration of stalling events, P.1201.1 only, no initial loading)

P.1201.2 inputs & outputs (1)





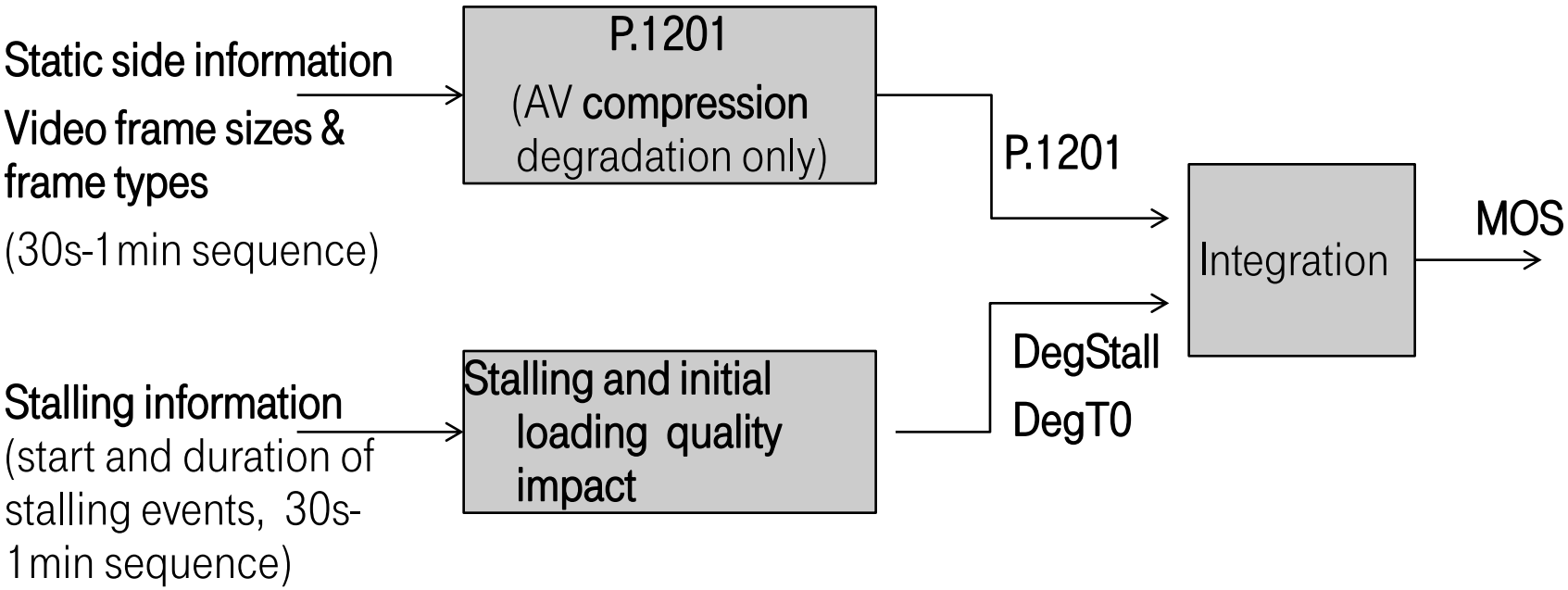
Qo_X: base quality (X≡**A**: audio, X≡**V**: video)

Icod$_X$: quality impact due to compression (**cod**ing) artefacts

Itra$_X$: quality impact due to **tra**nsmission error artefacts

(1) Garcia et al. *Parametric model for audiovisual quality assessment in IPTV: ITU-T Rec. P.1201.2*, MMSP 2013

P.1201-PD inputs & outputs



- P1201: quality impact due to audiovisual compression degradation (up to HVGA: P1201.1; for SD to HD1080: P1201.2)
- DegStall: quality impact due to stalling degradation[1]
- DegT0: quality impact due to initial loading degradation[2]

[1] Based on T. Hossfeld et al. (2013), Internet Video Delivery in YouTube: From Traffic Measurements to Quality of Experience. In: Data Traffic Monitoring and Analysis, E. Biersack et al. (Eds.), Springer,

[2] Based on T. Hossfeld et al. (2012), Initial delay vs. interruptions: between the devil and the deep blue sea, QoMEX 2012

P.1201 subjective test overview (1/3)

Training

- 15 (official) training databases

	Audio	Video	Audiovisual
P.1201.1	1	3	2
P.1201.2	1	5	3

Validation

- 24 validation databases

	Audio	Video	Audiovisual
P.1201.1	2	4	4
P.1201.2	2	8	4

- Validation criteria: RMSE (ITU-T P.1401)
- P1201 for progressive download: cross-validation procedure (5 audiovisual databases)

P.1201 subjective test overview (2/3)

- P1201.1 and P1201.2 experimental set-up
  - Absolute Categorical Rating (ACR, ITU-T P.800/P.910)
  - 8 contents per test, 10s-16s (if stalling) duration, TV contents
  - Simulated degradations: compression and transmission error s resulting in slicing and freezing degradations
  - 24-30 subjects per test
  - Standardized test rooms (ITU-R BT-500, ITU-T P.800/P.910) / professional equipment
- P1201-PD experimental set-up
  - Absolute Categorical Rating (ACR, ITU-T P.800/P.910)
  - 8 contents per test (no full design), 30s-45s/45s-1min contents, YouTube-like contents
  - Simulated degradations: compression and stalling (including initial loading)
  - 24-30 subjects per test
  - Standardized test rooms (ITU-R BT-500, ITU-T P.800/P.910) / professional equipment

P.1201 subjective test overview (3/3) – example P.1201.2

ITU-T P.NAMS	**Video**	**Audio**
Format	**Standard Definition** (SD, PAL: 720x576, NTSC: 720x480), **High Definition** (HD, 1920x1080, 1280x720)	44.1 kHz, 48 kHz, stereo presentation
Codec	**H.264** main & high profile	AAC-LC, HE-AAC, MPEG1-LII, AC3
Bitrate	0.5 to 30 Mbps	16 to 384 kbps
Packet Loss	0 to 2% Random & bursty	0 to 6% Random & bursty
Packet Loss Concealment	**Freezing with skipping , Slicing**	Codec built-in
Encoder settings	**1 slice/frame, 1slice/macroblock-row**	n.a.

P.1201 Models performance

r: Pearson's correlation coefficient

RMSE: Root-Mean-Square-Error (on 5-point scale)

Table 9 – Performance information for ITU-T P.1201.1 (LR)

	RMSE	Pearson correlation
Audiovisual	0.470 (based on 1166 samples)	0.852 (based on 1166 samples)
Video	0.535 (based on 1430 samples)	0.830 (based on 1430 samples)
Audio	0.351 (based on 690 samples)	0.941 (based on 690 samples)

Table 10 – Performance information for ITU-T P.1201.2 (HR)

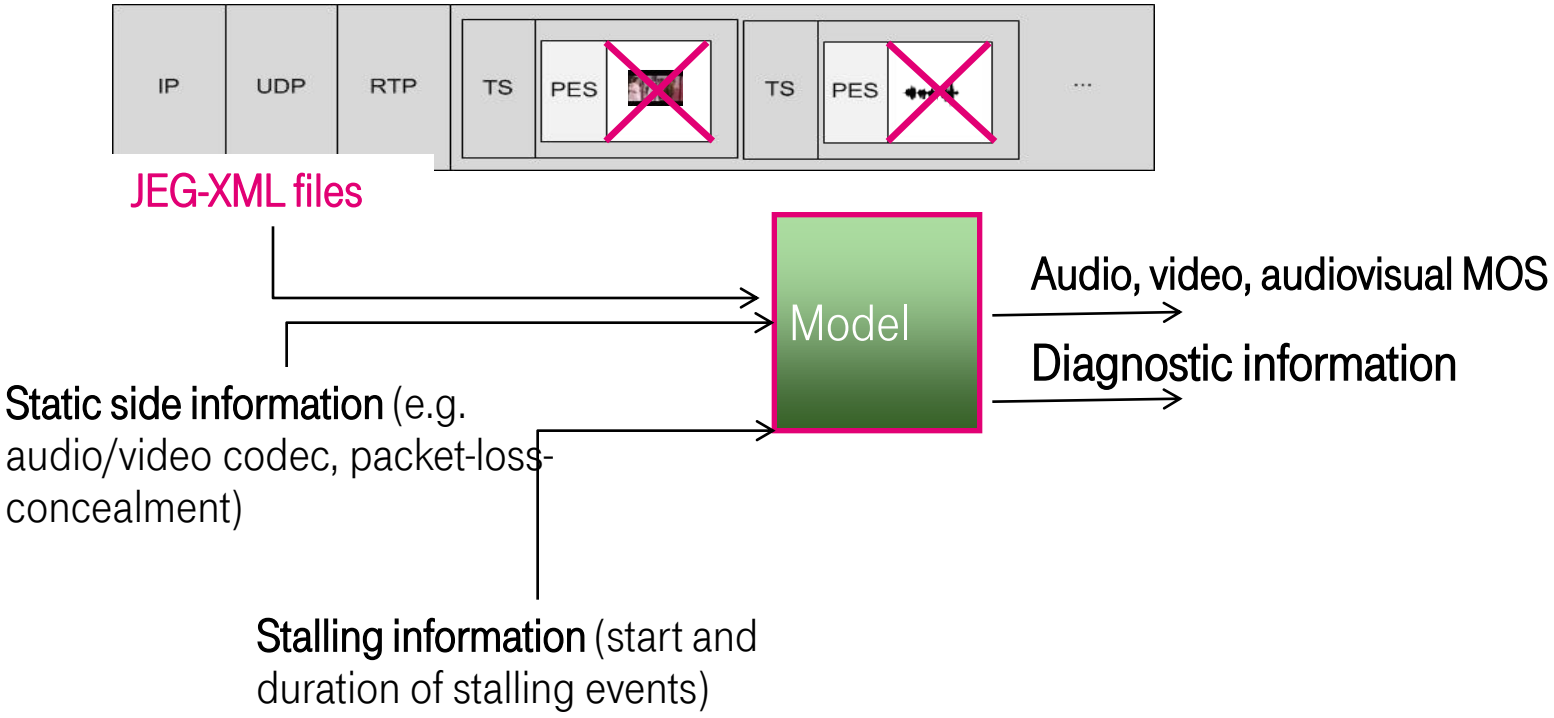
	RMSE	Pearson correlation
Audiovisual	0.435 (based on 3190 samples, PES and TS encrypted)	0.911 (based on 3190 samples, PES and TS encrypted)
Video	0.461 (based on 6138 samples, PES and TS encrypted)	0.902 (based on 6138 samples, PES and TS encrypted)
Audio	0.336 (based on 1360 samples, PES and TS encrypted)	0.949 (based on 1360 samples, PES and TS encrypted)

r	RMSE
0.89	0.44

Table 13 – P.1201-PD model performance results.

Outlook

- ITU-T P.NATS: quality models for adaptive streaming
- For JEG: share of P.1201.2 + P.1201-PD models
  - Sharing and usage procedures will be clarified
  - PCAP-parameter extraction not provided; The shared models will take as inputs: JEG-XML files, static side information, and stalling information

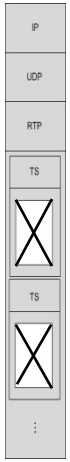


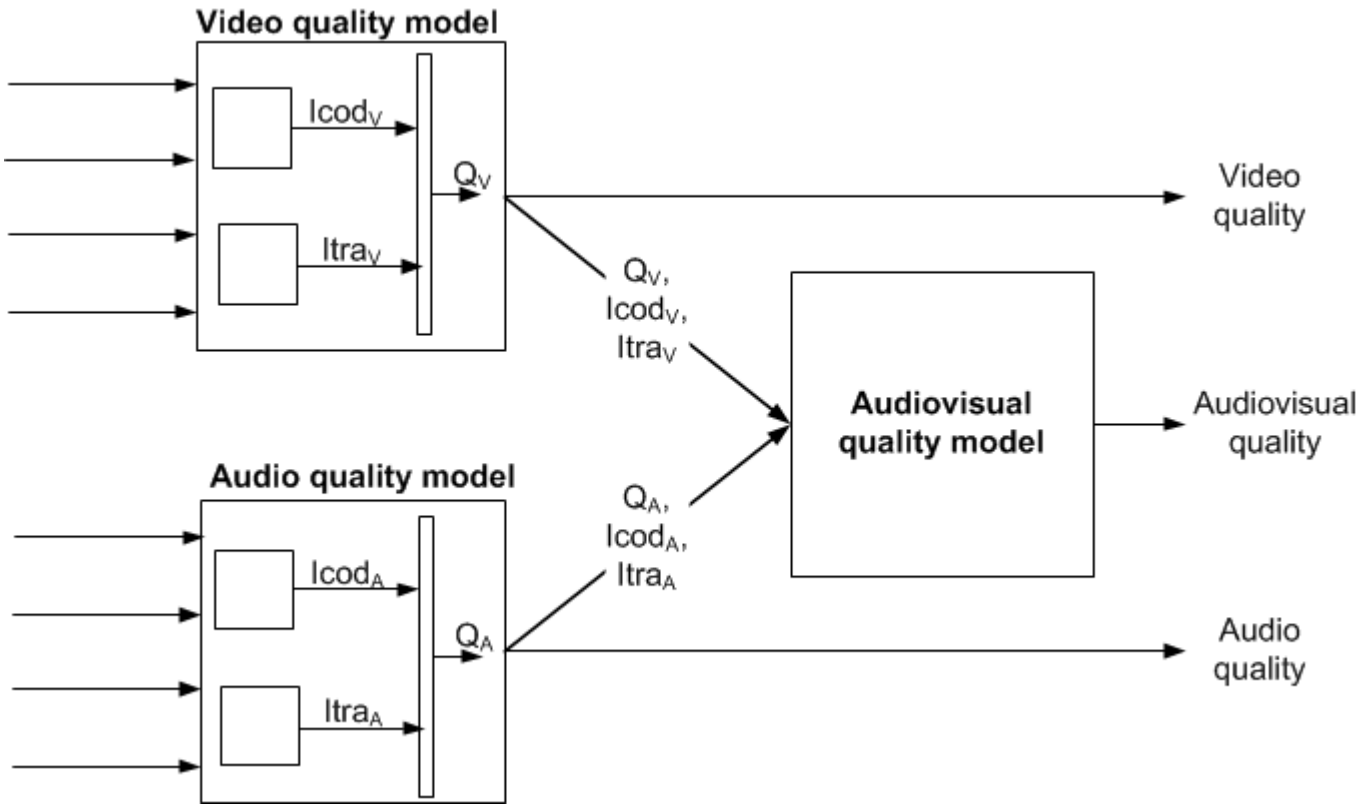
THANK YOU!

Back-up

- ITU-T P.1201.2 algorithm: slides 15-33
- ITUT P.1201 Appendix III (PNAMS-PD) algorithm: slides 34-35
- References: slide 36

P.1201.2 Model overview [1] (1/2)





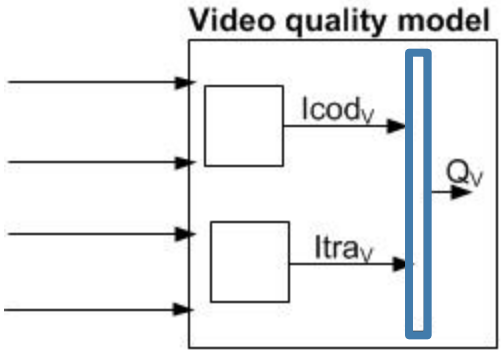
Qo_X: base quality (X≡**A**: audio, X≡**V**: video)

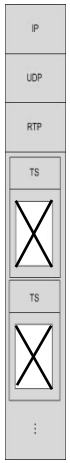
Icod$_X$: quality impact of compression (**cod**ing) artefacts

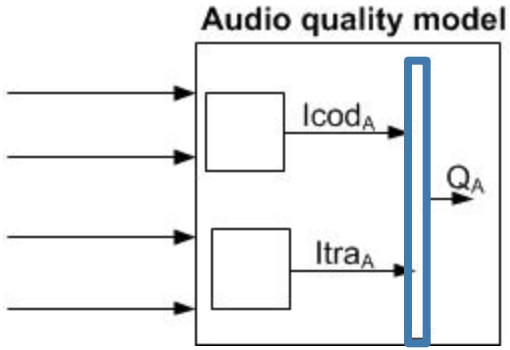
Itra$_X$: quality impact of **tra**nsmission error artefacts

(1) Garcia et al. *Parametric model for audiovisual quality assessment in IPTV: ITU-T Rec. P.1201.2*, MMSP 2013

P.1201.2 Model overview [1] (2/2)







$$Q_X = Qo_X - Icod_X - Itra_X$$

Impairment-factor-based approach [2, 3, 4]

Qo_X: base quality (X≡**A**: audio, X≡**V**: video)

Icod$_X$: quality impact of compression (**cod**ing) artefacts

Itra$_X$: quality impact of **tra**nsmission error artefacts

(1) Garcia et al. *Parametric model for audiovisual quality assessment in IPTV: ITU-T Rec. P.1201.2*, MMSP 2013

(2) ITU-T Rec. G.107: *the E-model, a computational model for use in the transmission planning*, 2005

(3) Garcia et al. *Parametric audio quality model for IPTV services – ITU-T P.1201.2 Audio*, QoMEX 2013

(4) Raake et al. *T-V-Model: Parameter-based prediction of IPTV quality*, ICASSP 2008

P.1201.2 Audiovisual Model

$$Qav = 0.3 \cdot Qav^{+} + 0.7 \cdot Qav^{*}$$

$$Qav^{+} = \alpha + \beta \cdot Qa + \gamma \cdot Qv + \mu \cdot Qa \cdot Qv \quad (1, 2)$$

$$\begin{aligned} Qav^{*} = Qoav &- c_{ac} \cdot IcodA - c_{vc} \cdot IcodV - c_{ac,vc} \cdot IcodA \cdot IcodV \\ &- c_{at} \cdot ItraA - c_{vt} \cdot ItraV - c_{at,vt} \cdot ItraA \cdot ItraV \\ &- c_{at,vc} \cdot ItraA \cdot IcodV - c_{ac,vt} \cdot IcodA \cdot ItraV \end{aligned} \quad (2)$$

- Modality dominance
- Interaction between audio and video quality
- **Impact of degradation type**

(1) Pinson et al. *Audiovisual quality components*, IEEE Signal Processing, 2011

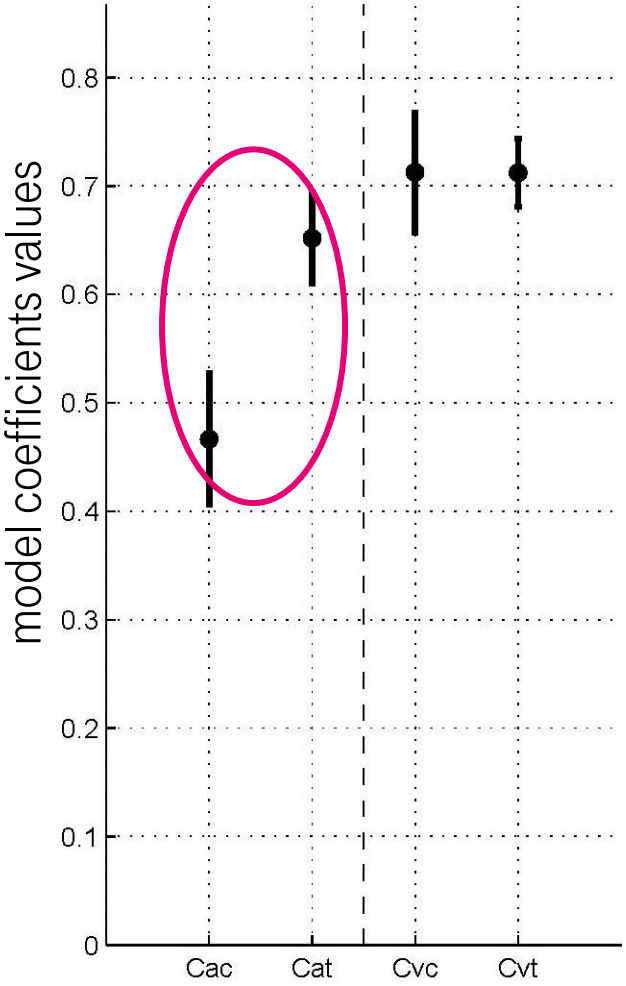
(2) Garcia et al. *Impairment-factor-based audiovisual quality model for IPTV...*, EURASIP Journal on Image and Video Processing, 2011.

P.1201.2 Audiovisual model

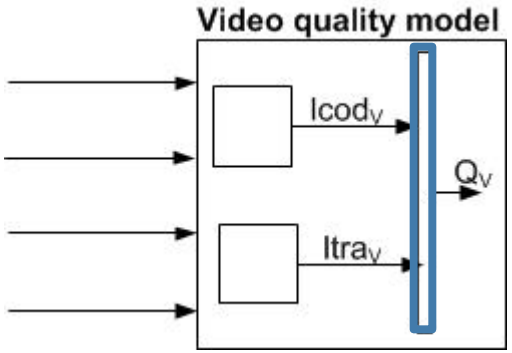
$$Qav^{*} = Qoav - c_{ac} \cdot IcodA - c_{vc} \cdot IcodV - c_{ac,vc} \cdot IcodA \cdot IcodV$$
$$- c_{at} \cdot ItraA - c_{vt} \cdot ItraV - c_{at,vt} \cdot ItraA \cdot ItraV$$
$$- c_{at,vc} \cdot ItraA \cdot IcodV - c_{ac,vt} \cdot IcodA \cdot ItraV$$

Impact of degradation type

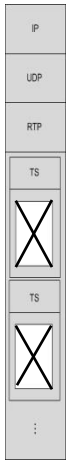
Attention attracted to transient audio degradations



P.1201.2 Video Model



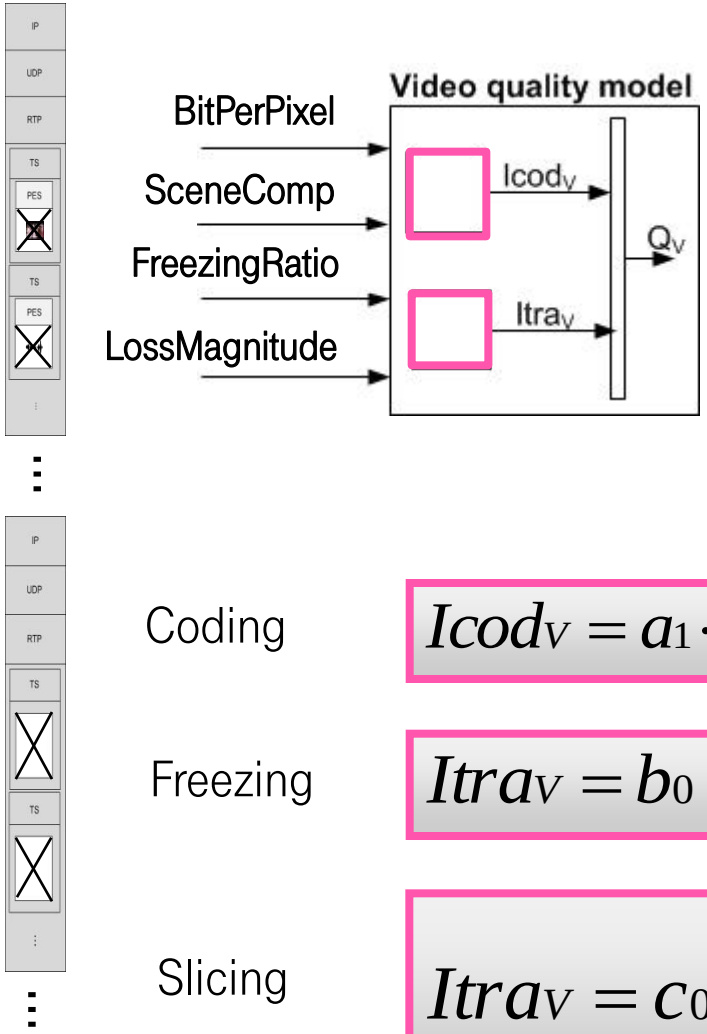
$$Q_V = Qo_V - Icod_V - Itra_V$$



Modeling objectives:

- Perceptually relevant packet-header based parameters
- Prevent for overtraining

P.1201.2 Video Model – Mapping to quality



$$Q_V = Qo_V - Icod_V - Itra_V$$

Coding

$$Icod_V = a_1 \cdot \exp(a_2 \cdot BitPerPixel) + a_3 \cdot SceneComp + a_4$$

Freezing

$$Itra_V = b_0 \cdot \log(b_1 \cdot FreezingRatio \cdot BitPerPixel + 1)$$

Slicing

$$Itra_V = c_0 \cdot \log(c_1 \cdot \frac{LossMagnitude}{Icod_{Vn}} + 1)$$

P.1201.2 Audio model

Base model

$$Q_A = 100 - Icod_A - Itra_A$$

Compression artefacts

$$Icod_A = a_1 \cdot \exp(a_2 \cdot bitrate) + a_3$$

bitrate: audio bitrate (kbps)

a2<0

Transmission error artefacts

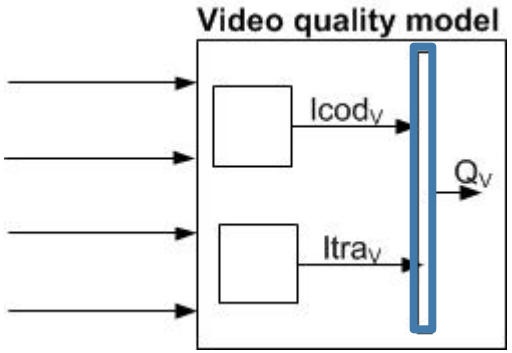
$$Itra_A = (b_0 - Icod_A) \cdot \frac{Pfl}{b_1 \cdot \mu + b_2 + Pfl}$$

Pfl: Percentage audio frame loss (%)

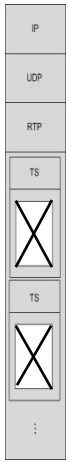
μ: burstiness, average number of consecutively lost audio frames

- Coefficient values depend on the audio codec
- Covered codecs: AAC-LC, HE-AAC, MP2, AC3, MP3

P.1201.2 Video Model



$$Q_V = Qo_V - Icod_V - Itra_V$$



Modeling objectives:

- Perceptually relevant packet-header based parameters
- Prevent for overtraining

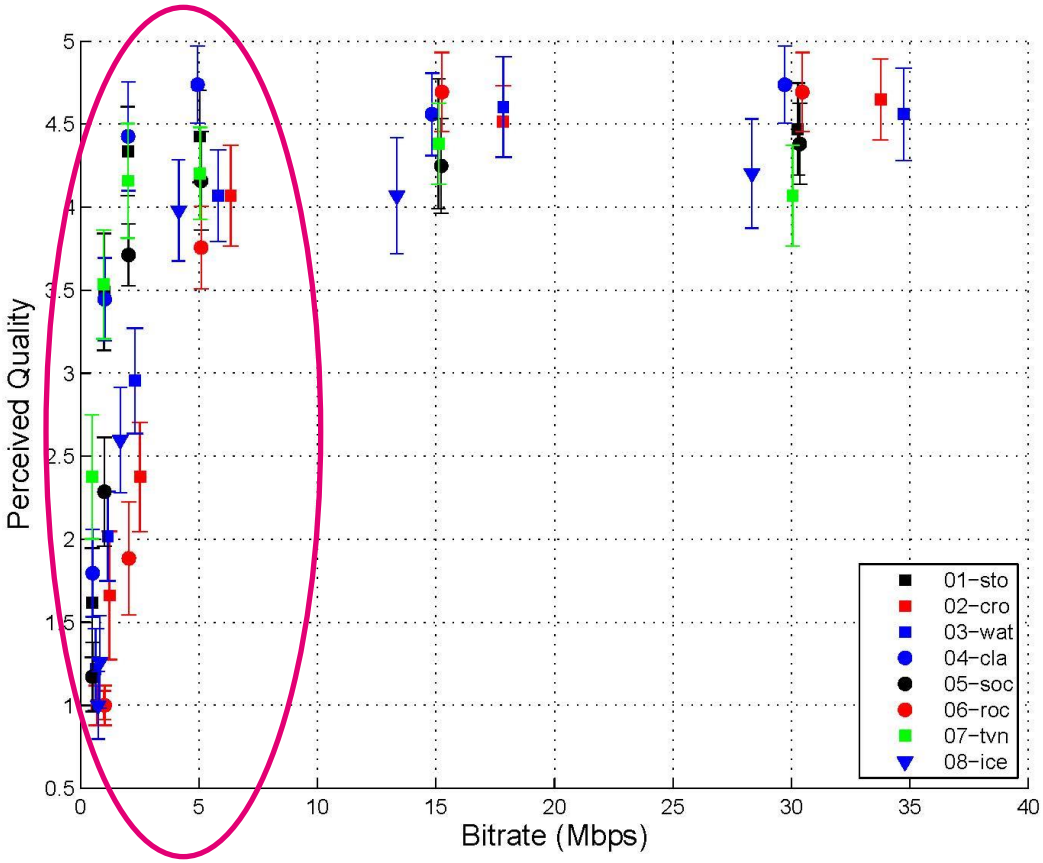
P.1201.2 Video Model – Compression artefacts

$$BitPerPixel = \frac{Bitrate \cdot 10^6}{nx \cdot fr}$$

bitrate: video bitrate (Mbps)[1,2]

nx: number of pixels per frame

fr: frame rate (fps)

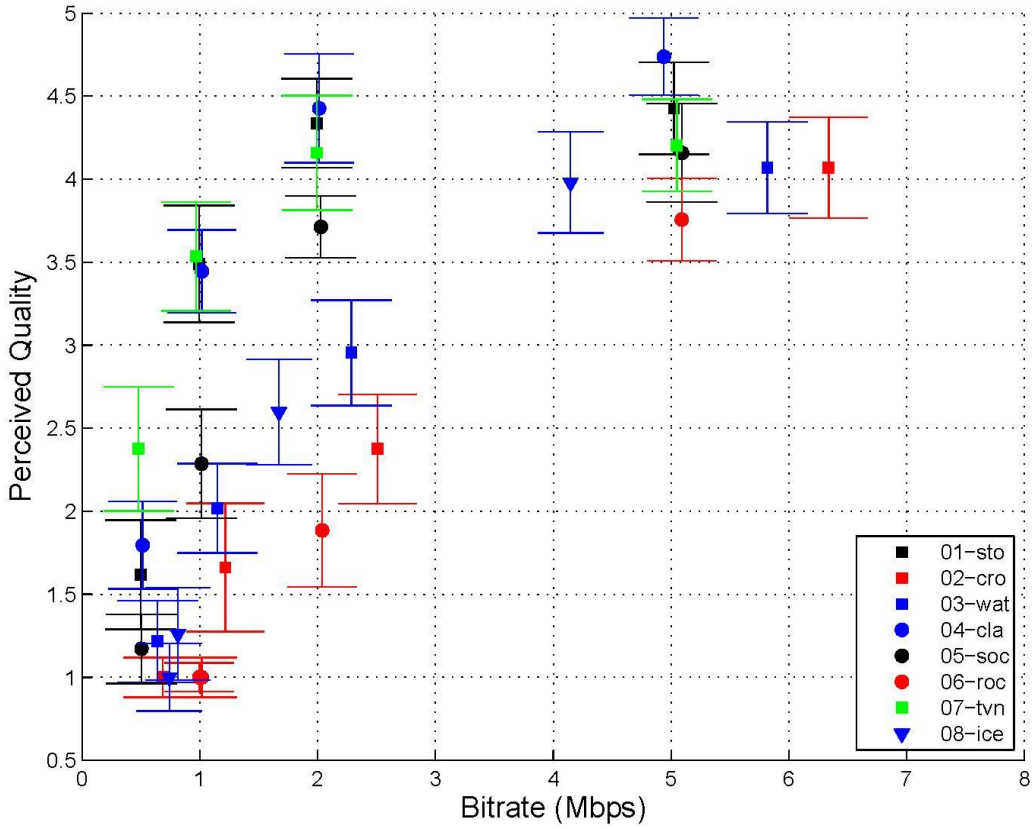


Content dependency

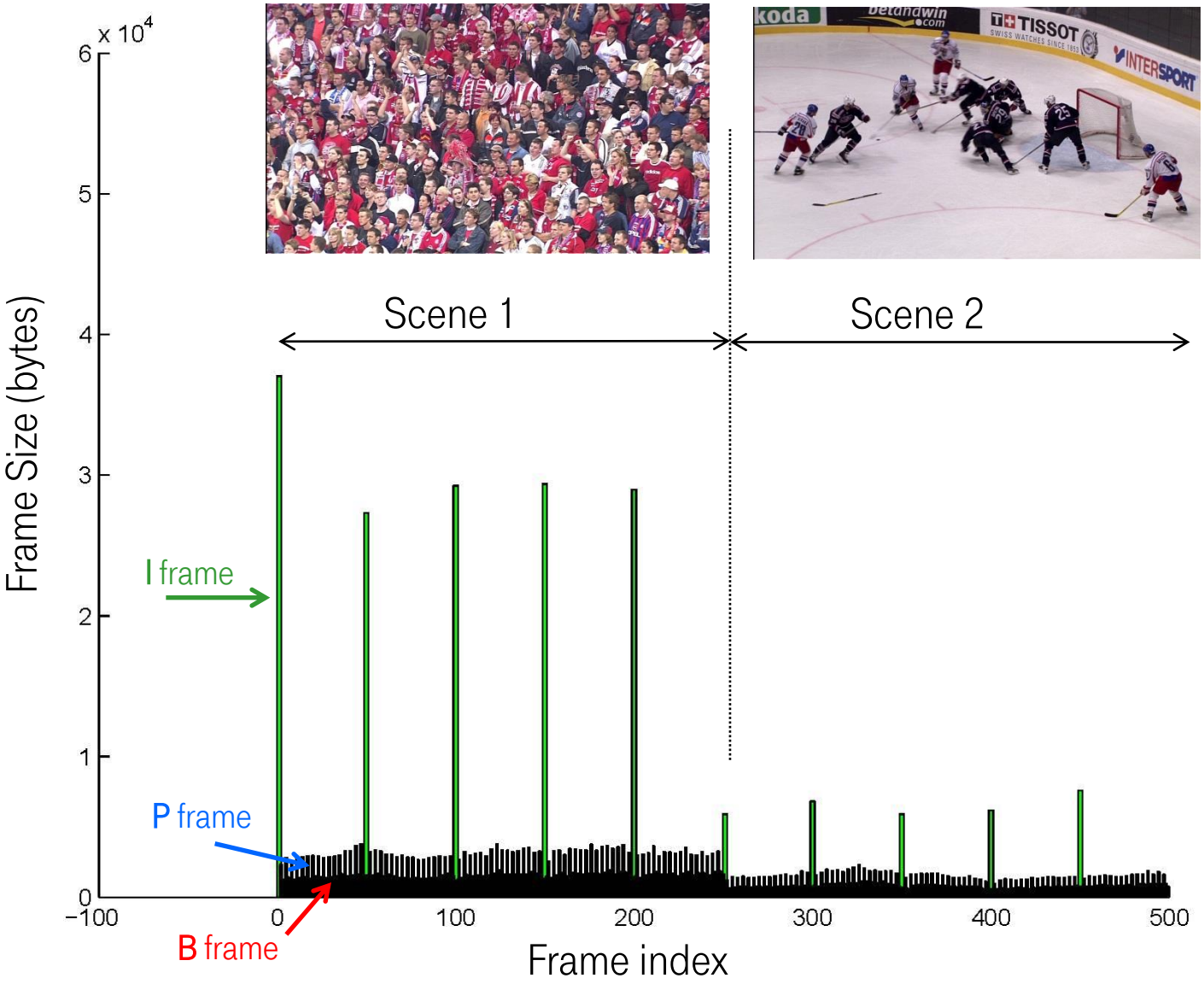
(1) Verscheure et al. *User-oriented QoS analysis in MPEG-2 video delivery*, Real-Time Imaging, 1999

(2) Yamagishi et al. *Parametric packet-layer model for monitoring video quality of IPTV services,* IEEE ICC, 2008

P.1201.2 Video Model – Compression artefacts



P.1201.2 Video Model – Compression artifacts



Low bitrate, two scenes, two spatio-temporal complexities

P.1201.2 Video Model – Compression artefacts

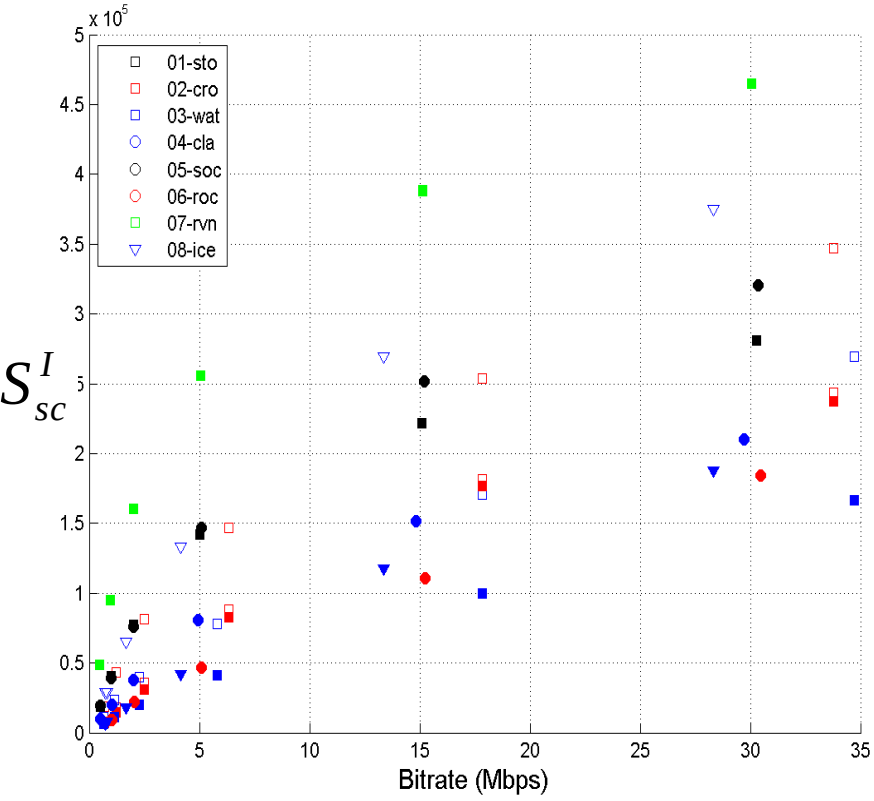
$$SceneComp = \frac{\sum_{sc=1}^{Z} w_{sc} \cdot N_{sc}}{\sum_{sc=1}^{Z} S_{sc}^{I} \, w_{sc} \cdot N_{sc}} \cdot \frac{nx \cdot fr}{1000}$$

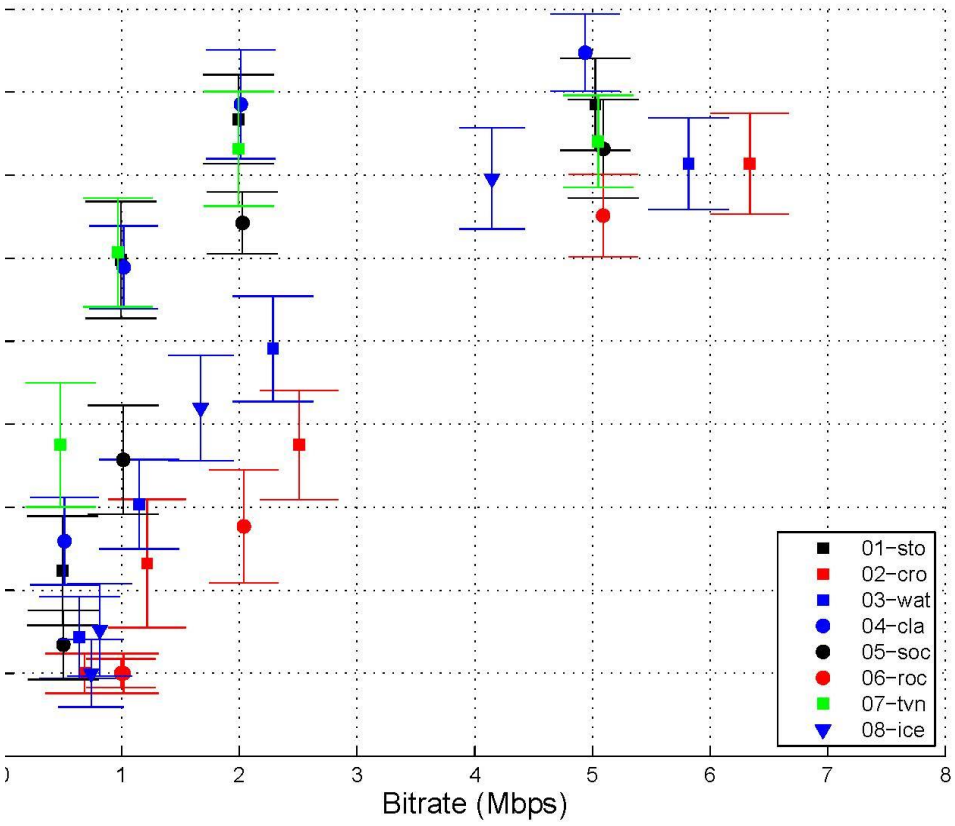
S_{sc}^{I} : average I frame sizes for scene sc

w_{sc} : higher weight on min S_{sc}^{I}

nx : number of pixels per frame

fr : frame rate





P.1201.2 Video Model – Transmission error artifacts

Typical packet-header based parameters [1,2,3]:

- Packet-loss-rate
- Burstiness
- Packet-loss-frequency

Do not capture:

- Loss duration (freezing & slicing)
- Spatial extent of the loss (slicing)
- Visibility of the loss (content dependency, slicing)

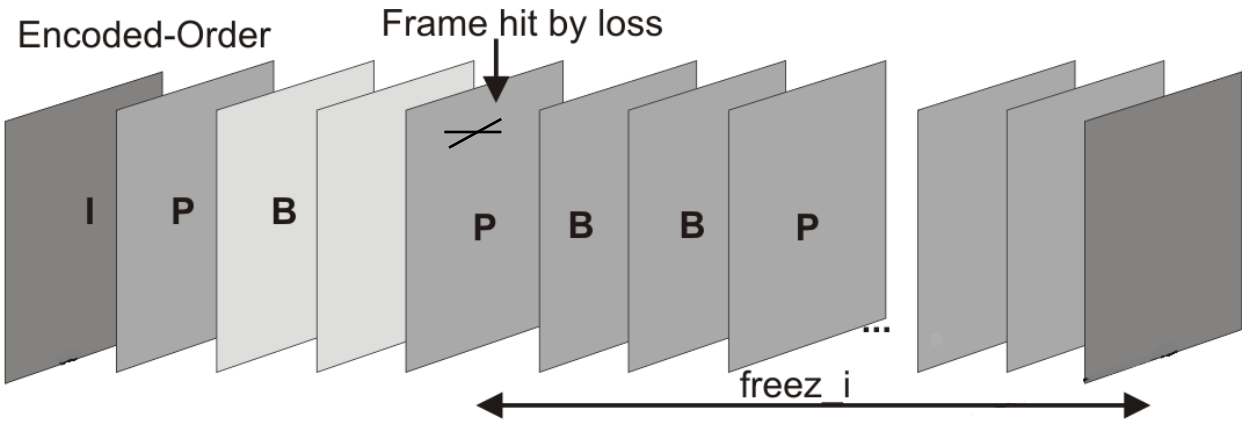
[1] Verscheure et al. *User-oriented QoS analysis in MPEG-2 video delivery*, Real-Time Imaging, 1999

[2] Yamagishi et al. *Parametric packet-layer model for monitoring video quality of IPTV services,* IEEE ICC, 2008

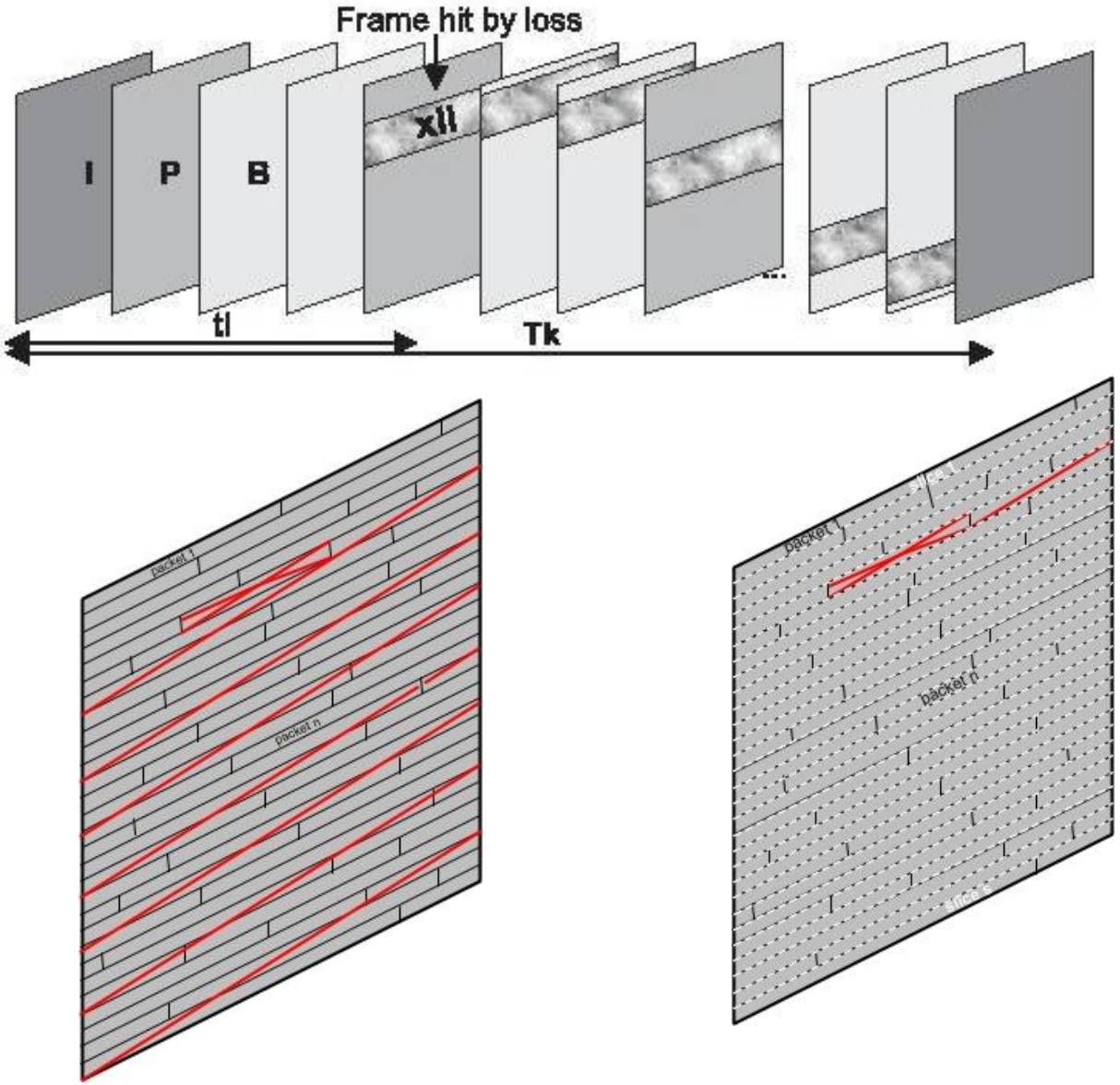
[3] You et al. *Packet loss pattern and parametric video quality model for IPTV*, ICIS, 2009

P.1201.2 Video Model - Freezing

$$FreezingRatio = \frac{\sum_{i=1}^{E} freez_i}{TotalNumberFrames}$$



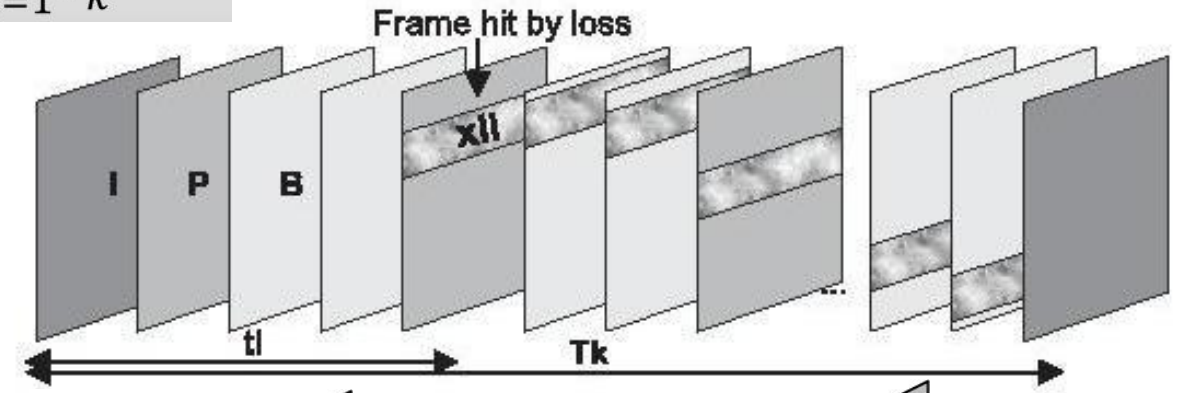
Typical IPTV degradations – Slicing

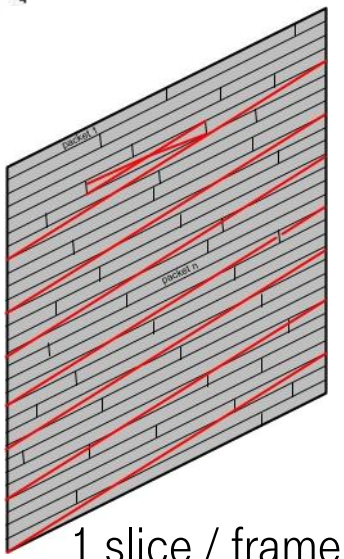


P.1201.2 Video Model - Slicing

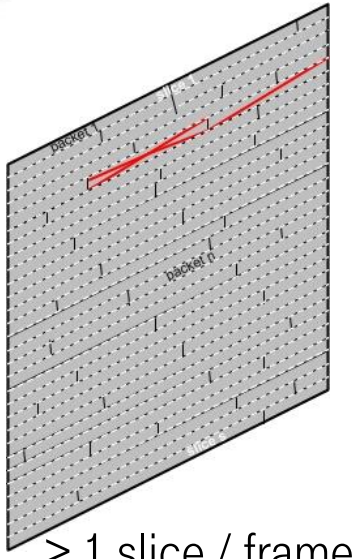
$$LossMagnitude = \frac{\sum_{k=1}^{G} c_k \cdot xl_k \cdot T_k}{\sum_{k=1}^{G} T_k}$$

$$xl_k = \frac{\sum_{i=1}^{N} xl_i \cdot (T_k - t_i)}{T_k}$$





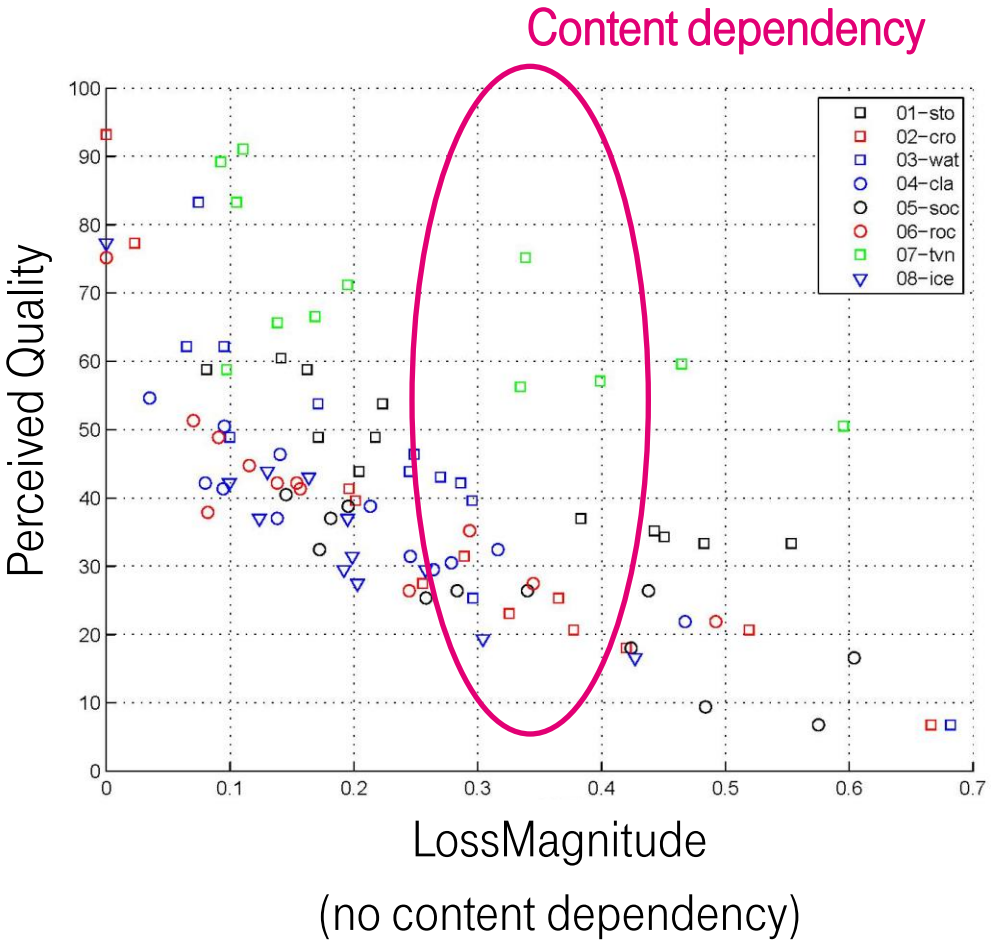
1 slice / frame



> 1 slice / frame

Garcia et al. *Frame-layer packet-based parametric video quality model for encrypted video in IPTV services* , QoMEX, 2011

P.1201.2 Video Model - Slicing



P.1201.2 Video Model - Slicing

$$LossMagnitude = \frac{\sum_{k=1}^{G} c_k \cdot xl_k \cdot T_k}{\sum_{k=1}^{G} T_k}$$

$$c_k = 0.5 \cdot \max\left(1 - \frac{S_k^b}{S_k^P}, 0\right) + 0.5 \cdot \min\left(2 \cdot \frac{S_k^{P+B+b}}{S_{sc}^I}, 1\right)$$

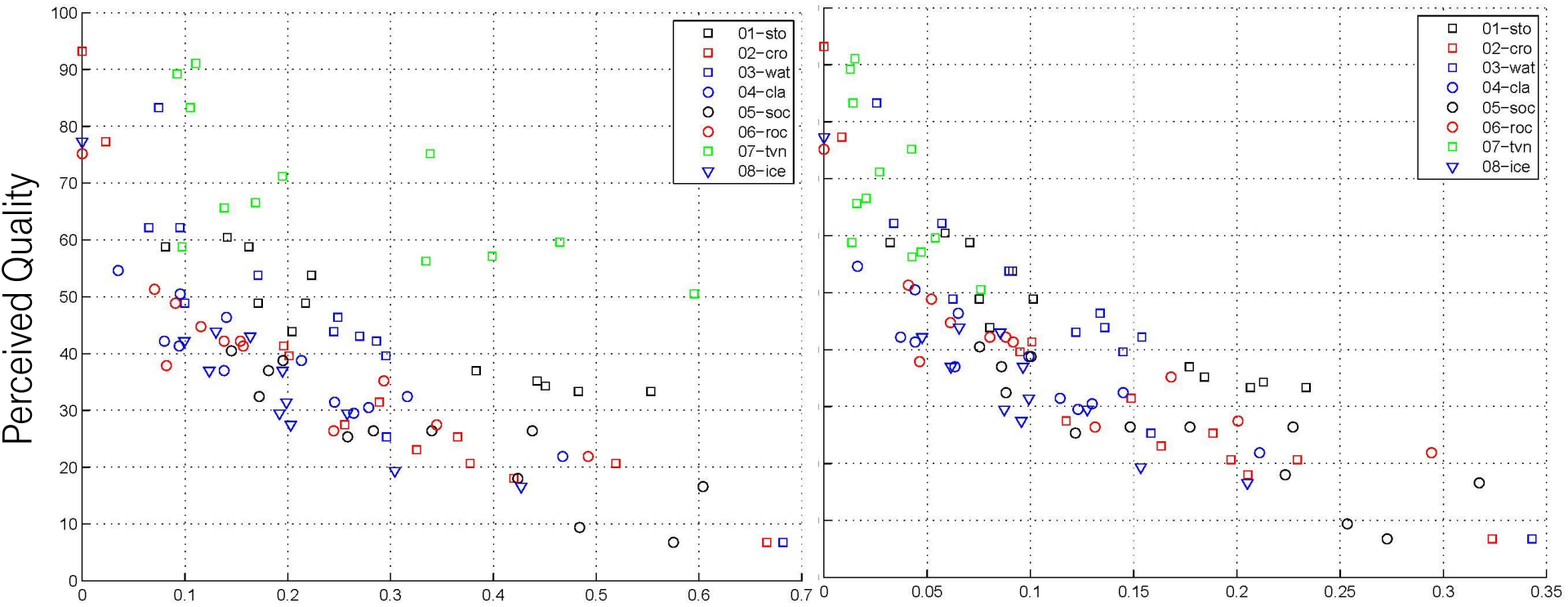
S_{sc}^{I} : average I frame sizes for scene *sc*

S_k^P : average P frame sizes for GOP k

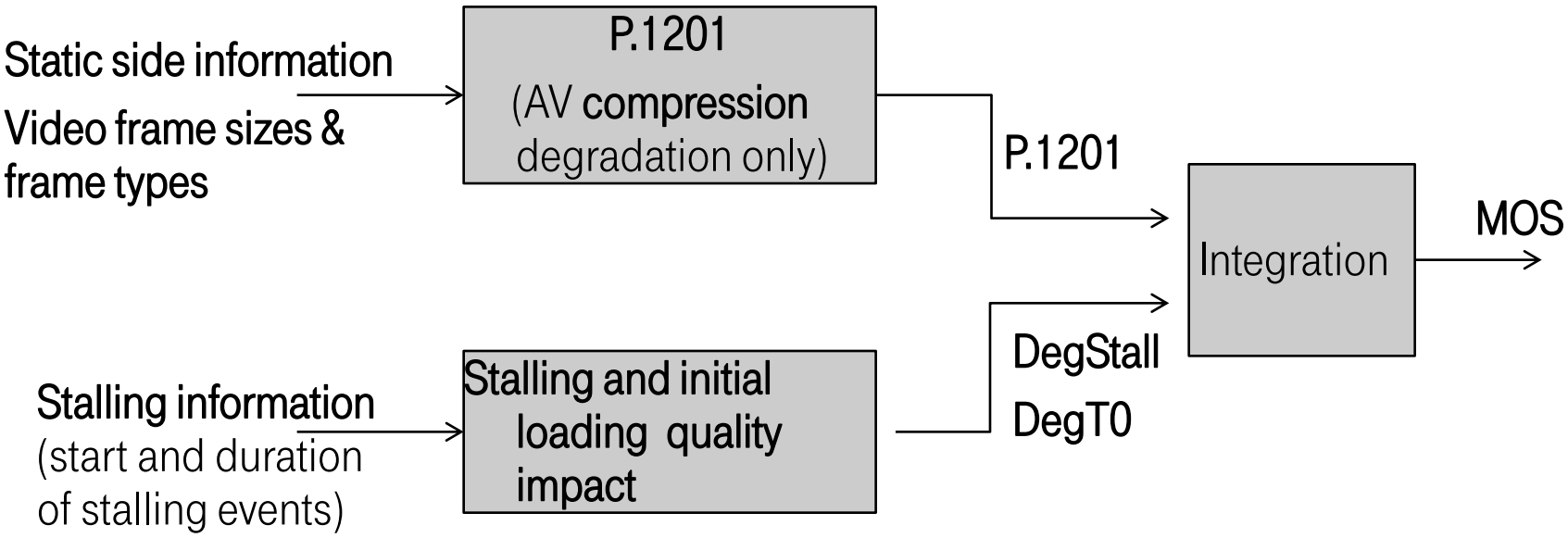
S_k^b : average non-reference B ("*b*") frame sizes for GOP k

S_k^{P+B+b} : average P, B and b frame sizes for GOP k

P.1201.2 Video Model - Slicing



P.1201 Appendix III (P.1201-PD / P.NAMS-PD) (1/2)



$$MOS = P1201 - DegStall - DegT0$$

- P1201: quality impact due to audiovisual compression degradation (up to HVGA: P1201.1; for SD to HD1080: P1201.2
- DegStall: quality impact due to stalling degradation (*)
- DegT0: quality impact due to initial loading degradation (**)

(*) Based on T. Hossfeld et al. (2013), Internet Video Delivery in YouTube: From Traffic Measurements to Quality of Experience. In: Data Traffic Monitoring and Analysis, E. Biersack et al. (Eds.), Springer,

(**) Based on T. Hossfeld et al. (2012), Initial delay vs. interruptions: between the devil and the deep blue sea, QoMEX 2012

P.1201 Appendix III (P.1201-PD / P.NAMS-PD) (2/2)

$$MOS = P1201 - DegStall - DegT0$$

$$DegStall = s_4 + s_1 \cdot \exp\left((s_2 \cdot AvgStallDur + s_3) \cdot NumStallEvents\right)$$

$$DegStall = \max(\min(DegStall, 4), 0)$$

$$DegT0 = \begin{cases} d_1 \cdot \lg(T0 + d_2), & if\ T0 > 1 - d2 \\ 0, & otherwise \end{cases}$$

$$DegT0 = \max(\min(DegT0, 4), 0)$$

- AvgStallDur: average duration (in s) of stalling events
- NumStallEvents: total number of stalling events
- T0: initial loading time (in s)

References

- A. Raake, J. Gustafsson, S. Argyropoulos, M.-N. Garcia, D. Lindegren, G. Heikkilä, M. Pettersson, P. List, and B. Feiten. IP-based mobile and fixed network audiovisual media services. In *IEEE Signal Processing*, volume 28. 2011.
- M.-N. Garcia, P. List, S. Argyropoulos, D. Lindegren, M. Pettersson, B. Feiten, J. Gustafsson, and A. Raake. Parametric model for audiovisual quality assessment in IPTV: ITU-T Rec. P.1201.2. In *Proc. of the 15th IEEE International Workshop on Multimedia Signal Processing (MMSP)*. 2013.
- T. Hoßfeld, R. Schatz, E. Biersack and L. Plissonneau, Internet Video Delivery in YouTube: From Traffic Measurements to Quality of Experience. In: Data Traffic Monitoring and Analysis, E. Biersack, C. Callegari, M. Matijasevic (Eds.), *Springer*. 2013
- T. Hoßfeld, S. Egger, R. Schatz, M. Fiedler, K. Masuch, C. Lorentzen, Initial delay vs. interruptions: between the devil and the deep blue sea, *QoMEX*. 2012
- M.-N. Garcia, A. Raake, and B. Feiten. Parametric audio quality model for IPTV services – ITU-T P.1201.2 audio. In *Proc. of the fith International Workshop on Quality of Multimedia Experience (QoMEX)*. 2013.
- A. Raake, M.N. Garcia, S.Moeller, J. Berger, F. Kling, P. List, J. Johann, and C. Heidemann. T-V-MODEL: Parameter-based prediction of IPTV quality. In *Proc. of International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*. 2008.
- M.H. Pinson, W. Ingram, and A. Webster. Audiovisual quality components. In *IEEE Signal Processing*, volume 28, pages 60–67. Nov. 2011.
- M.-N. Garcia, R. Schleicher, and A. Raake. Impairment-factor-based audiovisual quality model for IPTV: Influence of video resolution, degradation type, and content type. In *EURASIP Journal on Image and Video Processing*, volume 2011, pages 1–14. 2011.
- O. Verscheure e, P. Frossard, and M. Hamdi. User-oriented QoS analysis in MPEG-2 video delivery. In *Real-Time Imaging*. 1999.
- K. Yamagishi and T. Hayashi. Parametric packet-layer model for monitoring video quality of IPTV services. In *Proc. of IEEE International Conference on Communications (ICC)*. 2008.
- F. You, W. Zhang, and J. Xiao. Packet loss pattern and parametric video quality model for IPTV. In *Proc. of International Conference on Computer and Information Science (ICIS)*. 2009.
- M.-N. Garcia and A. Raake. Frame-layer packet-based parametric video quality model for encrypted video in iptv services. In *Proc. of International Workshop on Quality of Multimedia Experience (QoMEX)*. 2011.