NOKIA Bell Labs

UNIVERSIDAD
POLITÉCNICA
DE MADRID

Subjective Assessment of Adaptive Media Playout (AMP) for Video Streaming

[1]Pablo Pérez, [2]Narciso García, and [1]Álvaro Villegas

5-6-2019 – 11th International Conference on Quality of Multimedia Experience (QoMEX 2019)

[1]Nokia Bell Labs, [2]Universidad Politécnica de Madrid

QoMEX 2019
11th International Conference on Quality of Multimedia Experience

What is AMP?

…and why should I care about it?


WR 9.58
OR 9.69
0.0

Did you notice it?

NOKIA Bell Labs


WR 9.58
OR 9.69
0.0

What is Adaptive Media Playout?
…and why should I care about it?

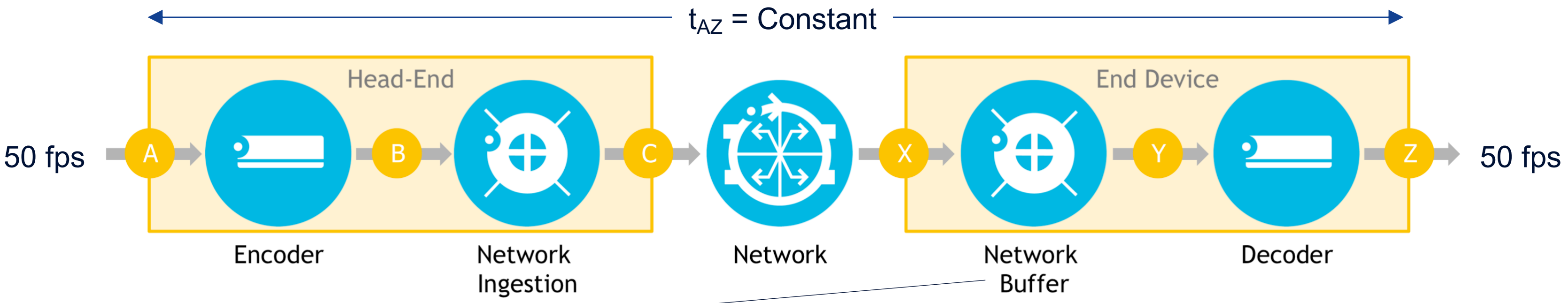
- What is AMP?
  - Dynamically changing playout speed at the video client
  - Without modifying audio pitch (Waveform Similarity Overlap-Add)

- Why should you care about it?

What is Adaptive Media Playout?

…and why should I care about it?

“In live video streaming, end-to-end delay must remain constant for the whole session”

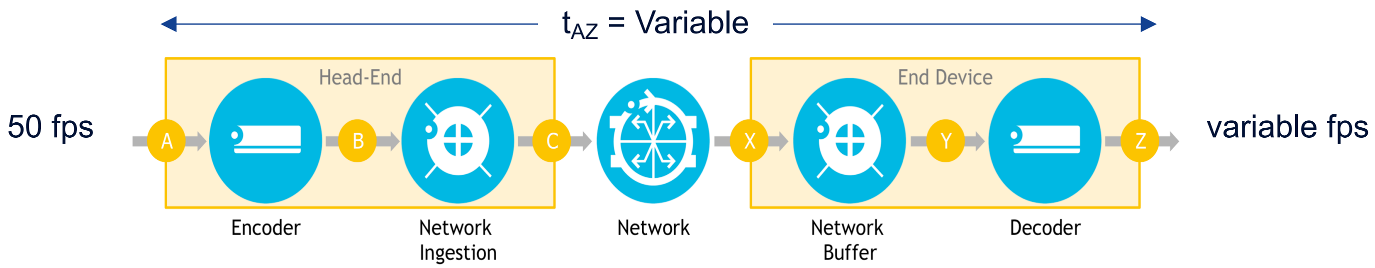


Buffer size (seconds) decided at the beginning

- Too low → underrun (stalling)
- Too high → high delay

What is Adaptive Media Playout?
…and why should I care about it?

- Changing playout speed ➔ changing end-to-end delay
  - E.g. synchronize two players viewing the same stream (IDMS)



- AMP is an interesting subjective assessment problem
  - Well-defined artifact, simple to generate
  - Good test bench for subjective assessment methodologies and models

What is Adaptive Media Playout?
…and why should I care about it?

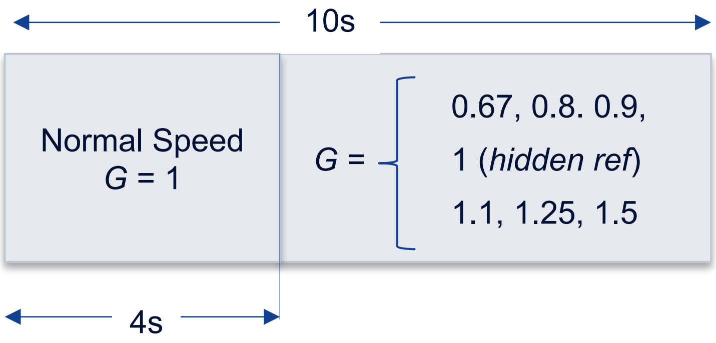
- …but there are very few papers characterizing the effect of AMP in QoE
  - Most prior art uses ad-hoc rules (“Up to 20% gain”) and symmetric cost models
  - [Rainer & Timmerer 2014] → Only one source content
  - [Mu *et al*. 2017] → Only speed increase

Target: Analyze subjective effect of AMP, including subject & content effect

Experimental Design & Methodology

Experimental Design
Selection of content

- 15 SRCs, 7 HRCs
- “Demanding, but not unduly so”
- Popular content (sports, well-known speakers, well-known movies…).
- 720p50, stereo audio



ID	Type	Name	Description
01	Sports	Sprint	Bolt wins 100m final run at Olympics
02	Sports	Goal	Real Madrid scores in football match
03	Sports	NBA	Kobe Bryant scores at NBA
04	Music	Radetzky	*Radetzky March* at New Year Concert
05	Music	Queen	*We Are The Champions* music video
06	Music	Sobral	Salvador Sobral at Eurovision final
07	Speech	PM	Prime Minister speech at Parliament
08	Speech	Show	Magic trick at TV show ‘El Hormiguero’
09	Speech	News	Matias Prats introducing news
10	Fiction	Crime	Parody of crime scene show
11	Fiction	Tiempo	*El Ministerio del Tiempo* TV show
12	Fiction	Galaxy	*Guardians of the Galaxy* animation
13	Action	Rogue	*Rogue One* space battle scene
14	Action	Clone	*Clone Wars* animation: light saber fight
15	Action	Wall-E	*Wall-E* animation: robots

Experimental Design
Methodology and population

- Tests on computer
  - 21” HD screen
  - Headphones
  - Mplayer (*scaletempo* plugin for AMP)
  - User scores after each PVS
  - Full randomization

- Degradation Category Rating (ITU-T P.910)

Normal Speed (G = 1)	Modified Speed (G = …)	Rating (DCR)

- 50 subjects (20 female, 30 male)

Subject Model

Jointly analyzing contribution to MOS of user/source

- We model subject score as a random process (similar to [Janowski & Pinson 2015])
  - Factor contributions are modeled as sum of independent gaussians
  - Main contribution: break PVS "ground truth": $\psi_j = \psi_{k,g} \approx \varphi_g + \Lambda_k$

Score for

- Subject i
- Source k
- Gain g

$$U_{i,k,g} = \underbrace{\overbrace{\varphi_g}^{\text{AMP score}}}_{} + \overbrace{\underbrace{\Delta_i}_{\text{subject bias}} + \underbrace{\upsilon_i X}_{\text{subject inconsistency}}}^{} + \underbrace{\overbrace{\Lambda_k}^{\text{content resilience}} + \overbrace{\rho_k Y}^{\text{content ambiguity}}}_{}$$

$$X, Y \sim \mathcal{N}(0,1)$$

Subject Model

Solving with MLE

- We compute variables by Maximum Likelihood Estimation (MLE) using Netflix Sureal framework [Li & Bampis 2017].

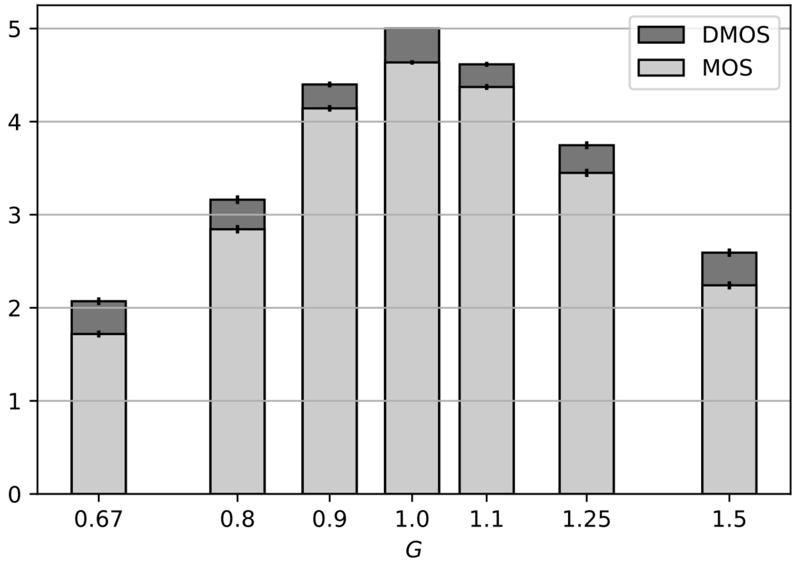
$$
\begin{aligned}
L(\theta) &= \log{(P)}(\{u_{i,k,h}\}|\theta) && (9)\\
&= \log{(P)}(\{u_{i,k,h}\}|\{\varphi_g\},\{\Lambda_k\},\{\rho_k\},\{\Delta_i\},\{\upsilon_i\}) && (10)\\
&= \sum_{i,k,g} -\frac{1}{2}\log{(\rho_k^2+\upsilon_i^2)} - \frac{1}{2}\frac{(u_{i,k,g}-\varphi_g-\Lambda_k-\Delta_i)^2}{\rho_k^2+\upsilon_i^2} && (11)
\end{aligned}
$$

Results

Effect of Rate Gain

Aggregate results

$$U_{i,k,g} = \boxed{\varphi_g} + \Delta_i + \upsilon_i X + \Lambda_k + \rho_k Y$$



- Safe limit: +/- 10%
- MOS (G) > MOS (1/G)
  - For any G > 1
- Same results using MLE / DMLE
- Simple cost model

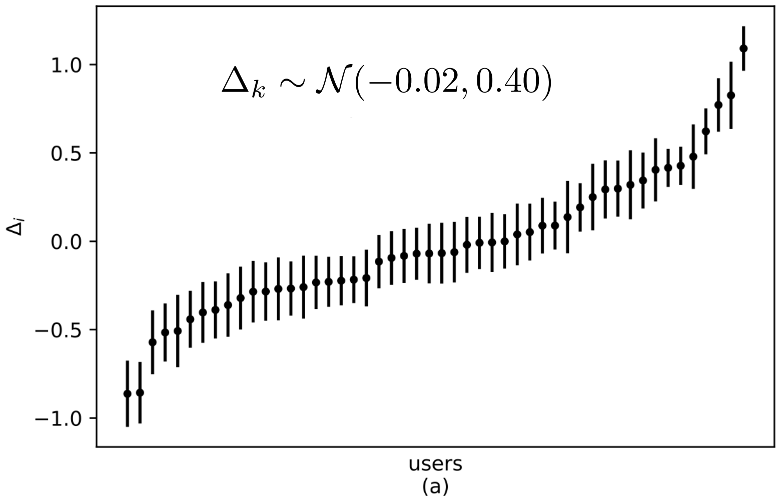
$$\mathrm{DMOS}(G) = \begin{cases} -4.1 + 9.1G, & \text{for } G \leq 1 \\ 9.9 - 4.9G, & \text{for } G > 1 \end{cases}$$

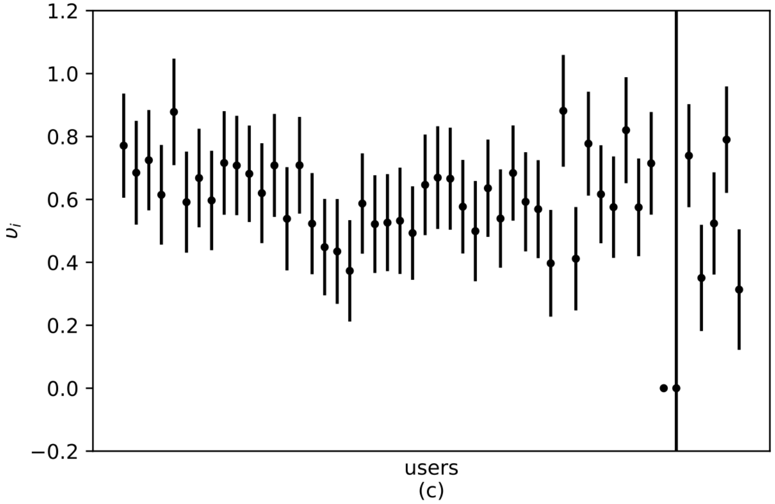
NOKIA Bell Labs

$$U_{i,k,g} = \varphi_g + \boxed{\Delta_i + \upsilon_i X} + \Lambda_k + \rho_k Y$$

Effect of Subject

Bias variability is higher than in (reported) video coding tests

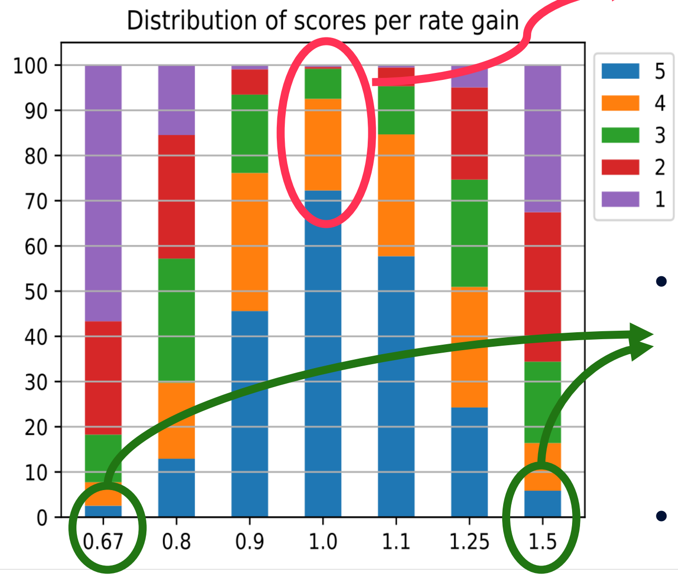




- Subject bias follows a normal distribution
  - Higher variance than [Janowski & Pinson 2015] (σ = 0.34).
- Subject inconsistency
  - Uncorrelated with bias

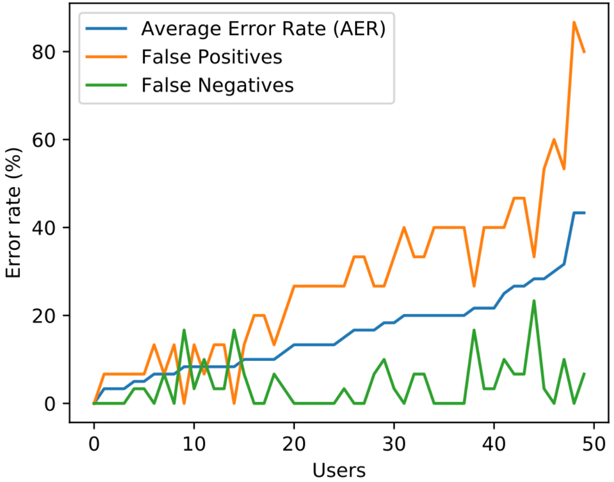
NOKIA Bell Labs

Effect of Subject
Some subjects are less reliable



- False Positives (~27%)
  - Impairment perceived, even when there is not!
- False Negatives (~5%)
  - No impairment perceived, even when it is strong!
- Errors unevenly distributed across subjects
  - Should we reject unreliable subjects?

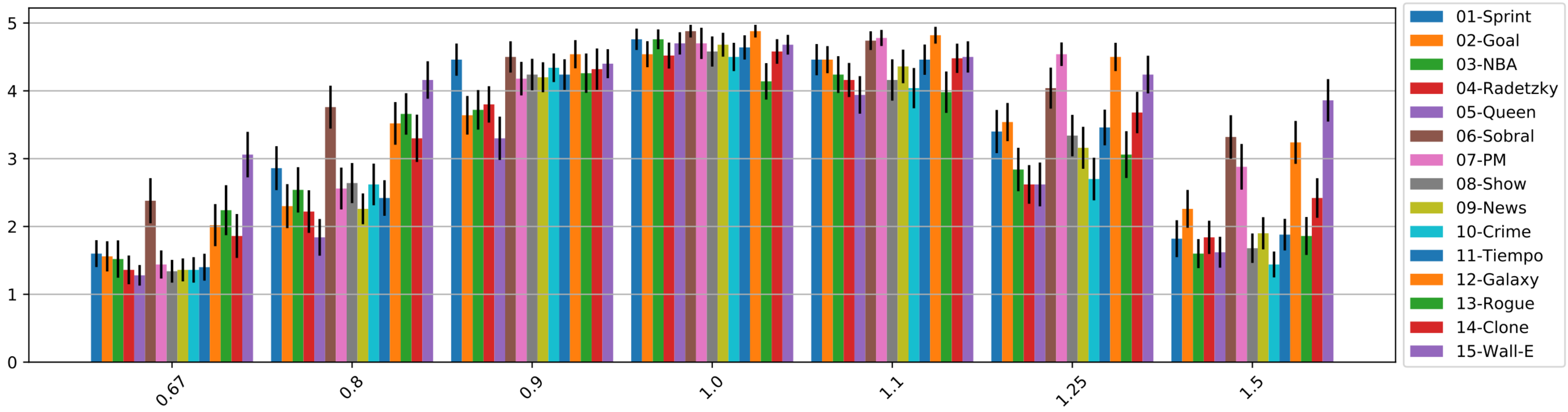
$$\mathrm{AER} = \frac{\mathrm{FalsePositives}(\%) + \mathrm{FalseNegatives}(\%)}{2}$$



NOKIA Bell Labs

Effect of Content

Significant variability between different SRCs

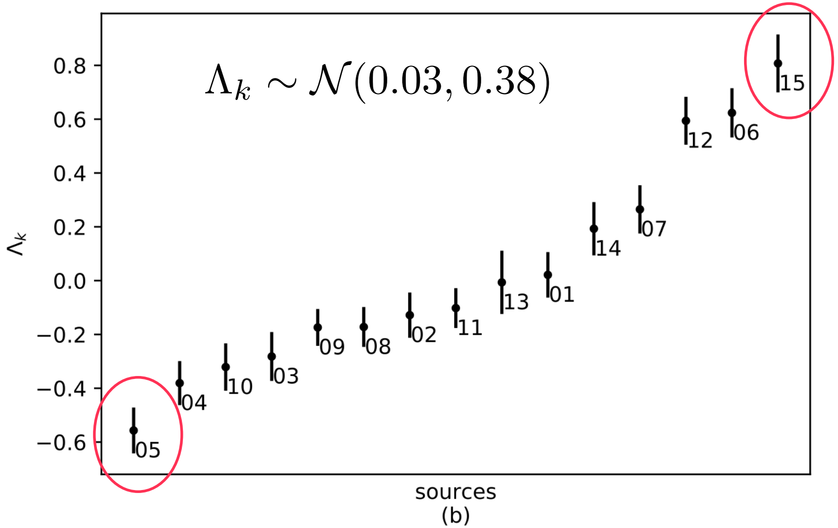


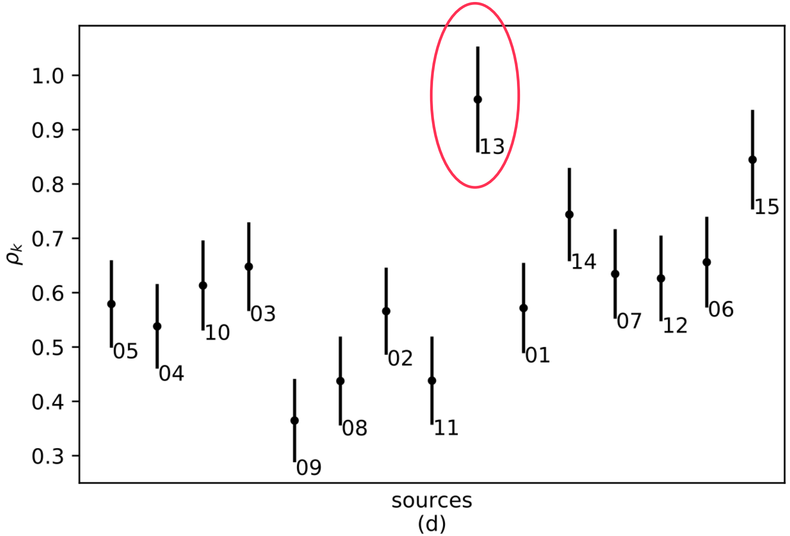
- Some sequences are more resilient to AMP

$$U_{i,k,g} = \varphi_g + \Delta_i + \upsilon_i X + \boxed{\Lambda_k + \rho_k Y}$$

Effect of Content

Strong variability of content resiliency and ambiguity





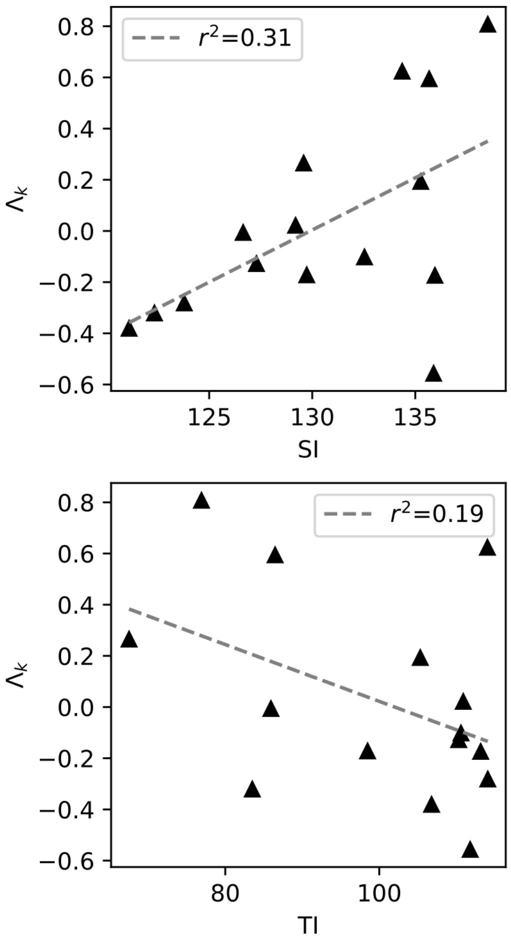
- *Content resilience* ~ normal distribution
  - 1.5 difference between highest (15: *Wall-E*) and lowest (05: *Queen*)
- Content ambiguity
  - Higher than in [Li & Bampis 2017]
  - Some sources (13: *Rogue One*) are extremely difficult to rate

NOKIA Bell Labs

Effect of Content

Content Analysis

- Significant difference from content to content
- Qualitatively
  - Best responses: animation (12, 14, 15), melodic music (06)
  - Worst responses: rhythmic music (04, 05)
  - Difficult to rate: action scenes (13-15)
- Quantitatively
  - No simple relationship with “trivial” video parameters
  - Weak correlation with SI/TI



Conclusions

Conclusions
Wrap Up

1. We have performed the most complete subjective test for AMP quality so far

2. We have provided practical guidelines for AMP implementation
   - “Rule of thumb”: 10% rate variation max
   - Slower speed is worse than higher speed

3. We have build a scoring model considering HRC and SRC fully separately
   - Useful for subject and content characterization
   - Could be used for other artifacts (e.g. compression)

4. We have characterized (qualitatively) *content resilience* to AMP
   - Quantitative characterization is not trivial: simple features (e.g. TI/SI) do not work

NOKIA Bell Labs

Subjective Assessment of Adaptive Media Playout (AMP) for Video Streaming

Questions?

5G PERFECTA

MINISTERIO
DE CIENCIA, INNOVACIÓN
Y UNIVERSIDADES

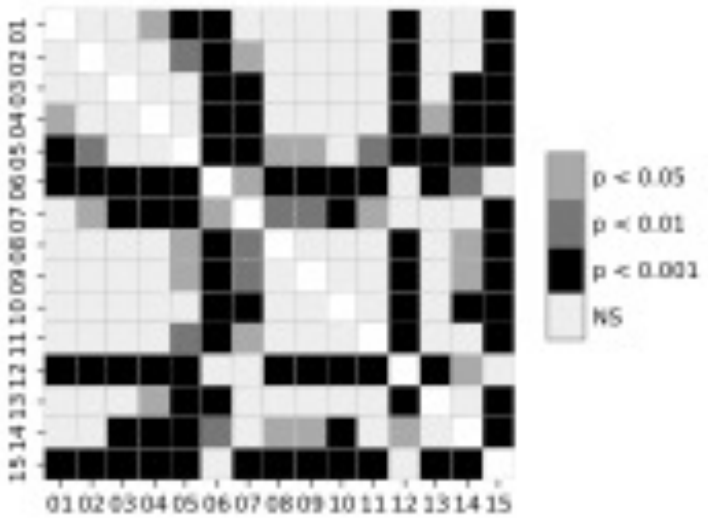
CDTI Centro para el Desarrollo Tecnológico Industrial

Statistical analysis
ANOVA and Tukey HSD

TABLE II
ANOVA TABLE FOR THE SCORES

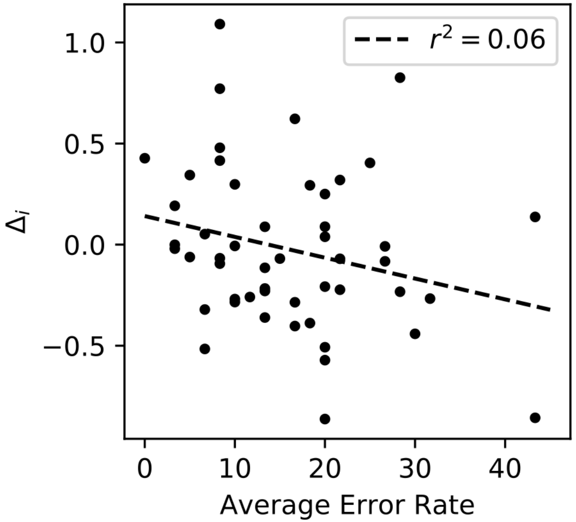
	SS	df	F	$P(>F)$	η^2	ω^2
C(G)	5612	6	1120	0.000***	0.50	0.50
C(SRC)	795	14	68	0.000***	0.07	0.07
C(G):C(SRC)	553	840	7.9	0.000***	0.05	0.04
Residual	4298	5145	-	-	-	-

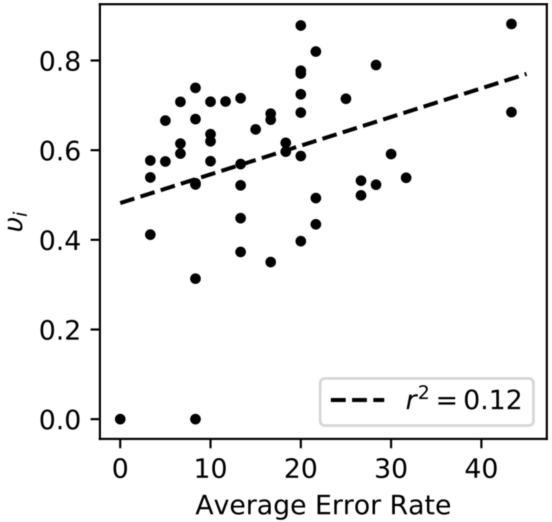
*** $p < 0.001$



NOKIA Bell Labs

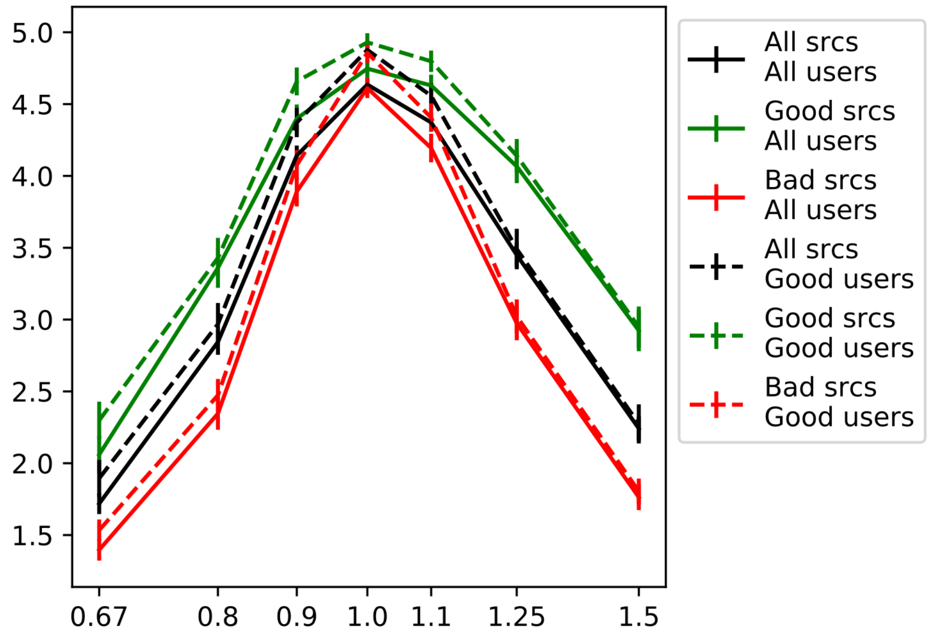
Relationship of AER vs Bias / Uncertainty





Effect of content and subjects

Subset of subject / contents



Comparison with prior art

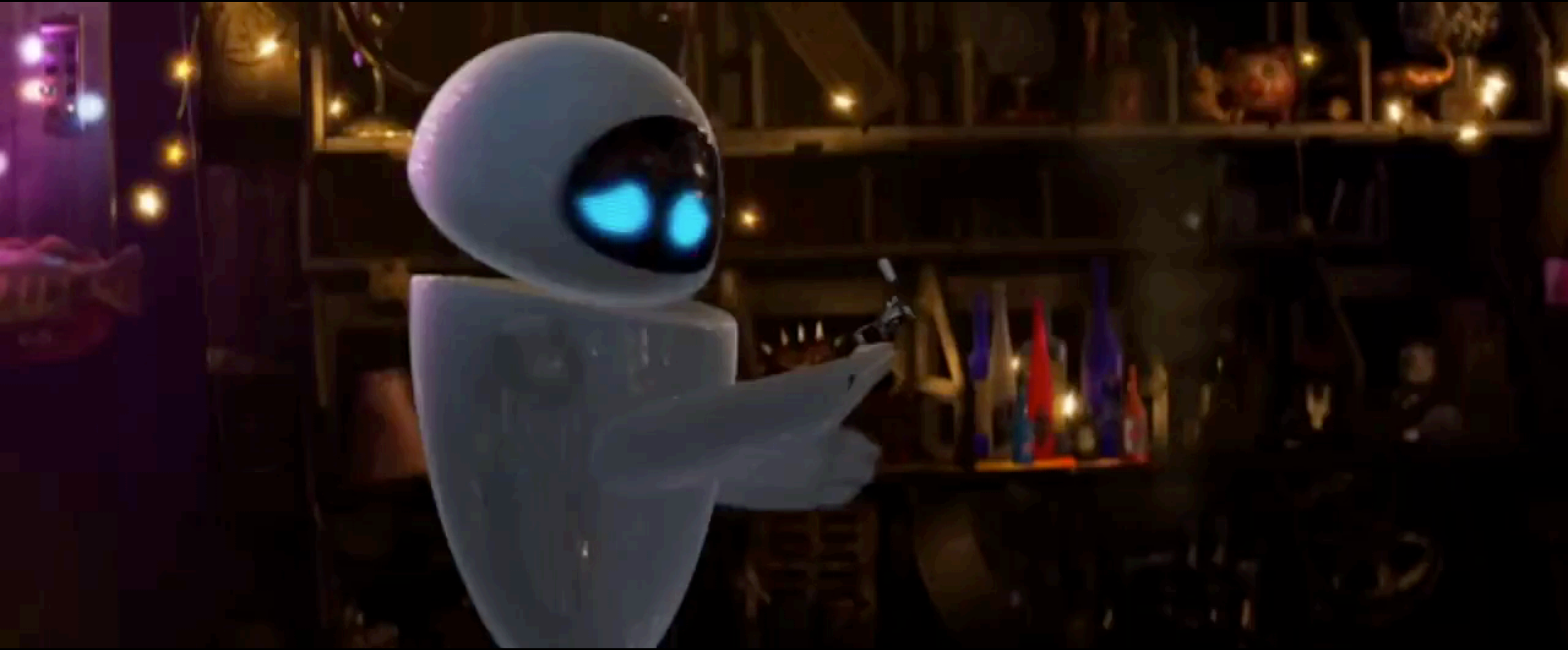
COMPARISON OF MODELS

G	MOS	DMOS	MLE	DMLE	[16]	[17]
0.67	1.72 ± 0.07	2.07 ± 0.08	1.67	2.04	4.51	-
0.80	2.84 ± 0.09	3.16 ± 0.09	2.78	3.11	4.81	-
0.90	4.14 ± 0.07	4.40 ± 0.06	4.17	4.39	4.94	-
1.00	4.64 ± 0.05	5.00 ± 0.00	4.68	5.00	5.00	4.99
1.10	4.37 ± 0.06	4.62 ± 0.05	4.41	4.60	5.00	4.45
1.25	3.45 ± 0.09	3.75 ± 0.08	3.46	3.74	4.97	3.74
1.50	2.24 ± 0.08	2.59 ± 0.09	2.22	2.58	4.79	2.80

$\pm CI$ means 95% Confidence Interval. CI for MLE and DMLE is 0.06 $\forall G$.

[16] Rainer & Timmerer, 2014

[17] Mu *et al.*, 2017




Animation & Heroes Movie
EnterTainments