UNIVERSITY OF CAGLIARI
DIEE - Department of Electrical and Electronic Engineering

Observing the Users to Estimate the Perceived Quality: Challenges and Technologies

Prof. Luigi ATZORI

We rely on the Qualinet QoE definition

Qualinet White Paper, 2013 [Qualinet]

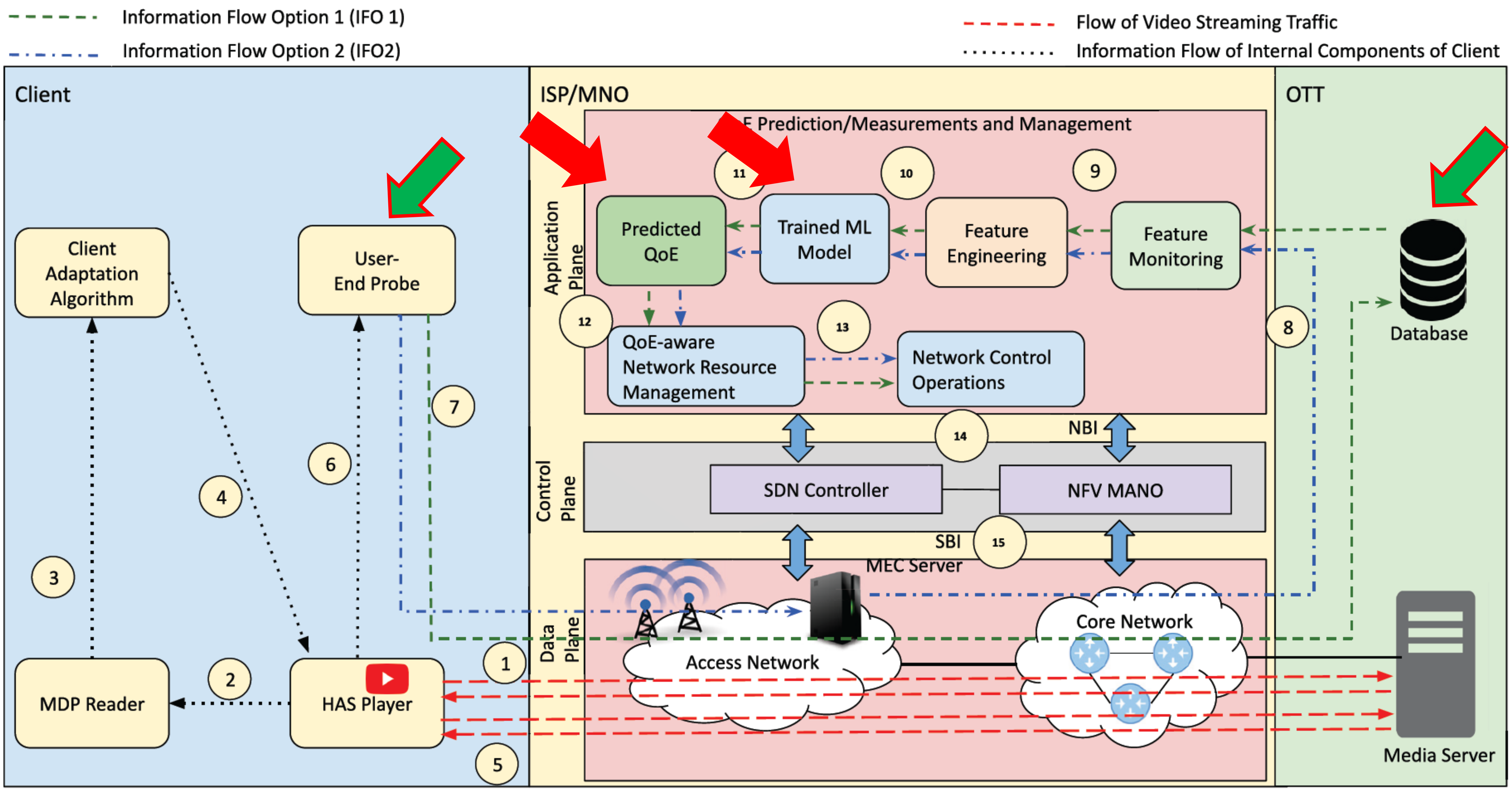
Output of the European Cost Action Qualinet

"The degree of delight or annoyance of the user of an application or service. It results from the fulfilment of his or her expectations with respect to the utility and/or enjoyment of the application or service in the light of the user's personality and current state"

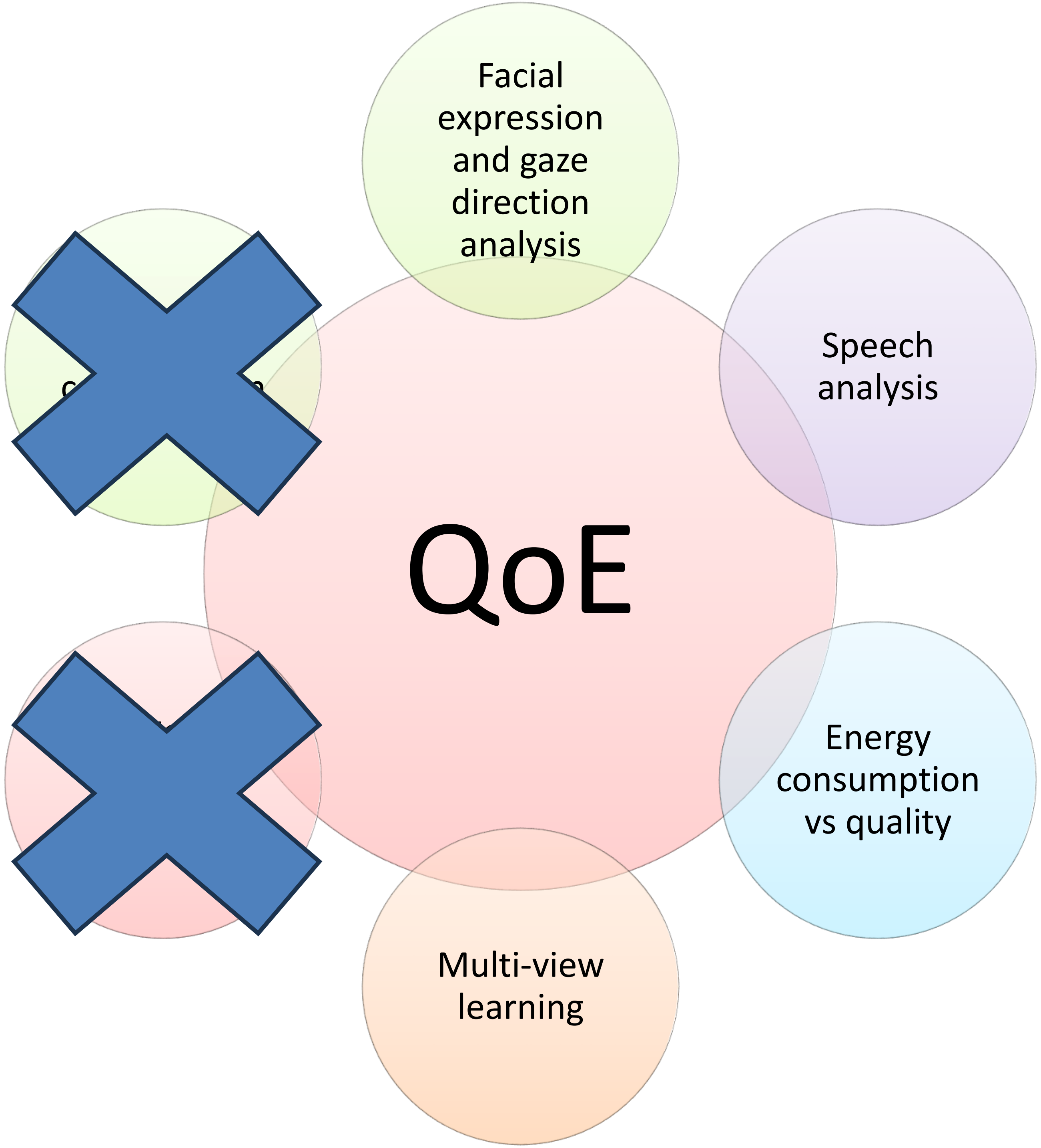
This definition has been adopted in 2016 by the International Telecommunication Union in Recommendation ITU-T P.10 updated.

QoE Influence Factors

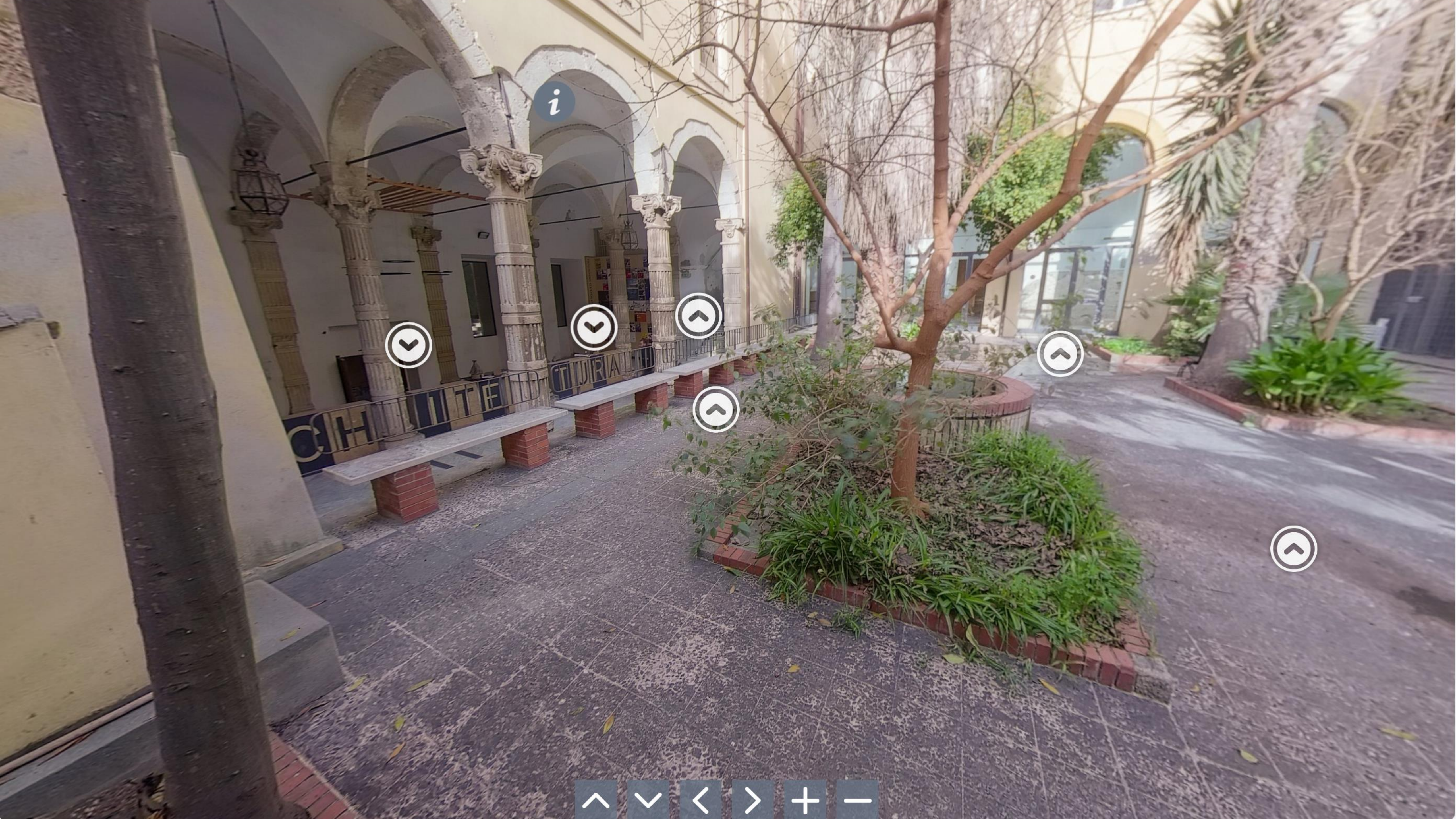
- **Influence Factor (IF)** are defined as any characteristic of a user, system, service, application, or context whose actual state or setting may have influence on the QoE
- Influence Factors must not be regarded as isolated as they may interrelate
- Influence Factors are grouped in three main dimensions:
  – Human IF
    - demographic and socio-economic background, the physical and mental constitution, or the user's emotional state
  – System IF
    - Content, media, network and device related factors
  – Context IF
    - Physical, cost, temporal, spatial, task



A. Ahmad, A. B. Mansoor, A. A. Barakabitze, A. Hines, L. Atzori and R. Walshe, "Supervised-learning-Based QoE Prediction of Video Streaming in Futu
Networks: A Tutorial with Comparative Study," in *IEEE Communications Magazine*, vol. 59, no. 11, pp. 88-94, November 2021


Facial expression and gaze direction analysis
Speech analysis
QoE
Energy consumption vs quality
Multi-view learning


luigi
visore
Svolgimento Esercitazione
LAB ENSA 5.4.13
Configurazione ACL IPv4 estesi
luigi sta condividendo il suo schermo
Svolgimento Esercitazione
LAB ENSA 5.4.13
Stanza
Meeting
Dettagli della stanza
Note, file e chat
Accesso ospite
nove
uno
Esci
luigi
1:26 PM



Video streaming

Video call

UNIVERSITY OF CAGLIARI
DIEE - Department of Electrical and Electronic Engineering

The need for models and the psychophysiological methods

Psychophysical assessment and modelling

Test conditions
- Application scenario
- Influence factors (delay, content type, spatial context)
- Subjects

Conduct the tests
- Laboratory
- In the wild
- Time-consuming
- ACR 5-level scale of rating often used

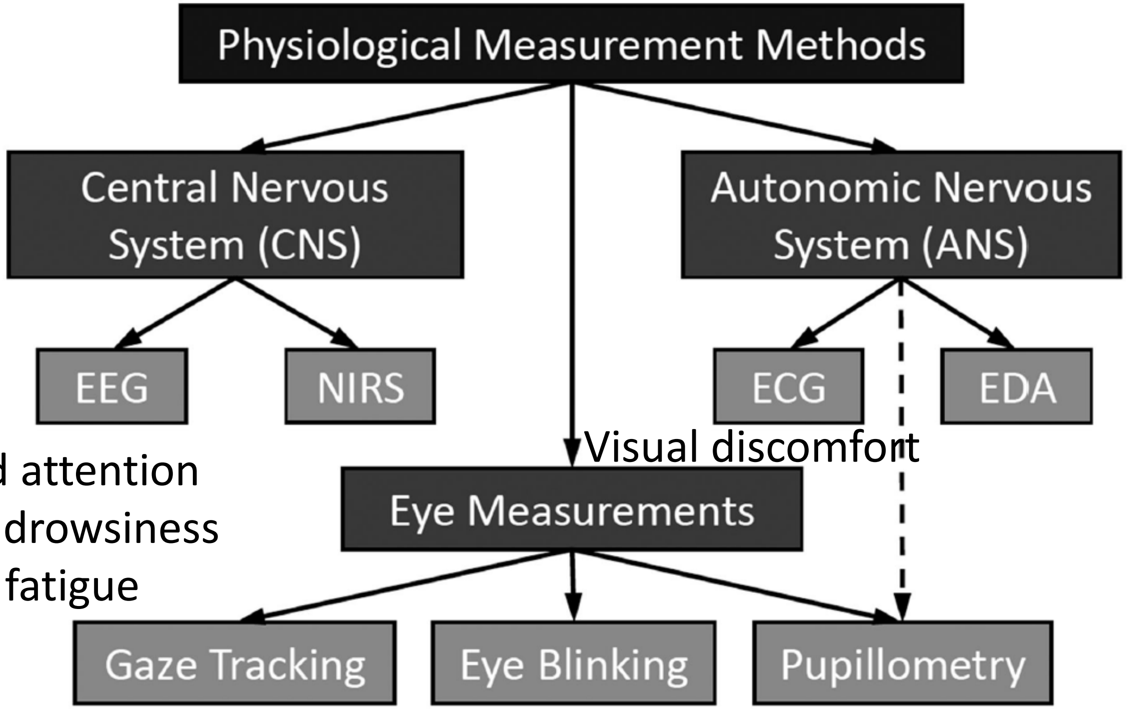
Define the model
- Parametric
- ML-based

Evaluate the performance
- Accuracy
- PCC
- MAE
- SROCC

- Quantitatively evaluate the relationship between physical stimuli and the conscious perceptions thereof
- User feedback is considered as the QoE ground-truth

Psychophysiology in QoE

- Psychophysiology: measurement of physiological signals and analysis of correlation with psychological processes



Decreased attention increased drowsiness alertness, fatigue

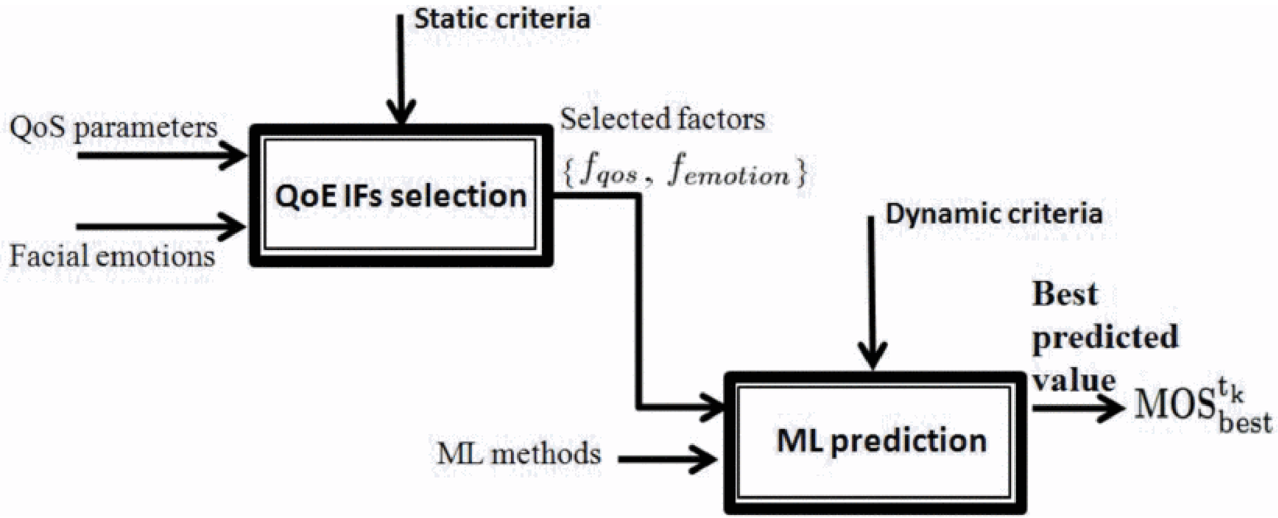
Visual attention

U. Engelke *et al.*, "Psychophysiology-based QoE assessment: A survey," *IEEE J. Sel. Topics Signal Process.*, vol. 11, no. 1, pp. 6–21, Feb. 2017.

- Implicit response: may overcome the problem of potentially misleading rating
- Relevant tests may reach ecological validity
- No substitution of psychophysical approaches but complement
- May be obstructive

Facial expression in QoE

- Facial expression is extensively used to estimate the user emotions
  - Mostly the six main emotions (fear, disgust, happiness, anger, surprise and sadness)
- The deviation of facial expression has been studied in AR applications [1]
  - Moderate positive correlation between the user feedback and micro facial expressions of disgust
- It has been used also to estimate the emotions and then QoE [2]
  - Using QoS parameters and facial emotions together obtained improvements in quality predictions



[1] E. of Lower Facial Micro Expressions as an Implicit QoE Metric for an Augmented Reality Procedure Assistance Application," *ISSC 2020*
[2] "An improved QHynes et al, "An Evaluation oE estimation method based on QoS and affective computing," 2018 International Symposium on Programming and Systems (ISPS), 2018

Research questions:

1) *is there any correlation between facial expression and perceived Quality of Experience?*

2) can we predict QoE directly from facial expression?

3) can the predictor be application independent?

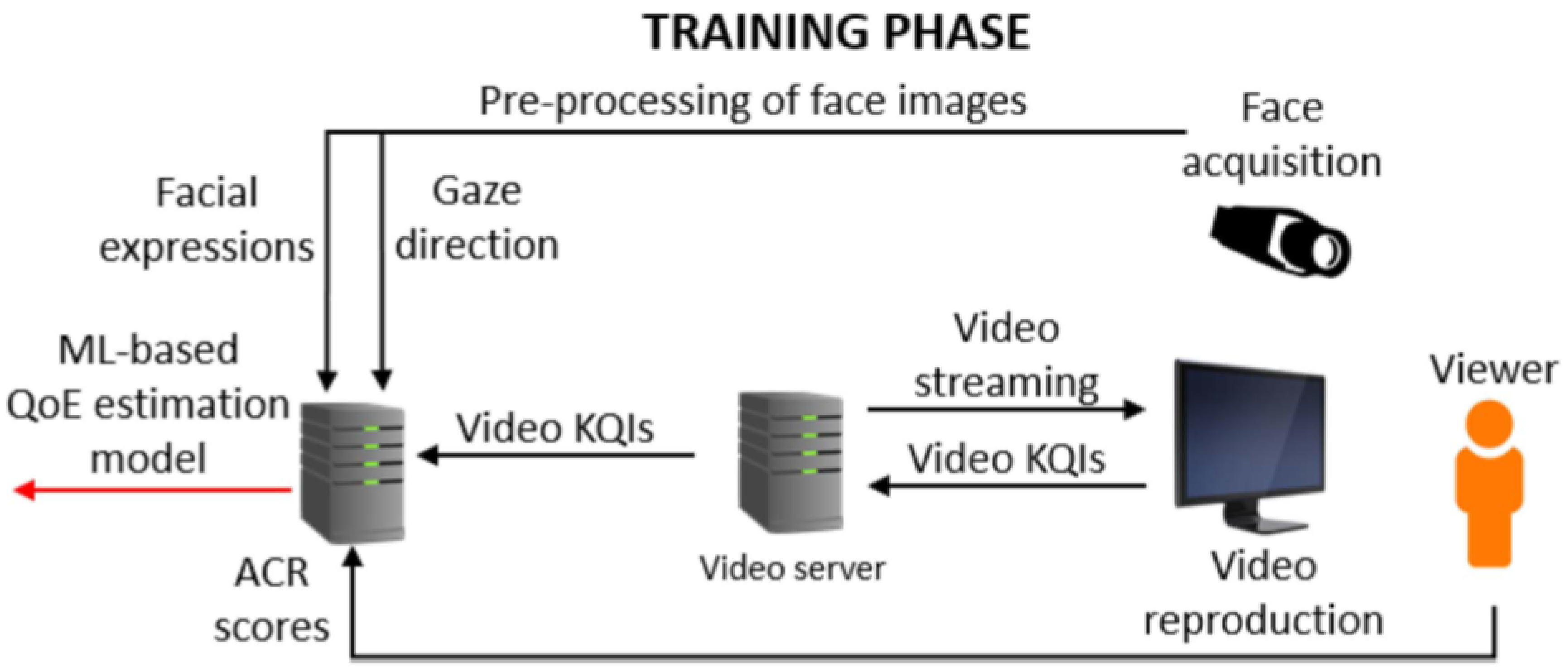
UNIVERSITY OF CAGLIARI
DIEE - Department of Electrical and Electronic Engineering

Towards the Prediction of the Quality of Experience from Facial Expression and Gaze Direction

Facial Action Units

- The Facial Action Coding System (FACS) separates the face into three parts (upper, middle and lower)
- Each of these parts is represented by Action Units (AUs), which identify specific muscle bands of the face
- The AUs provide information on the presence and intensity of muscle movement

Prediction of QoE in HAS video streaming sessions



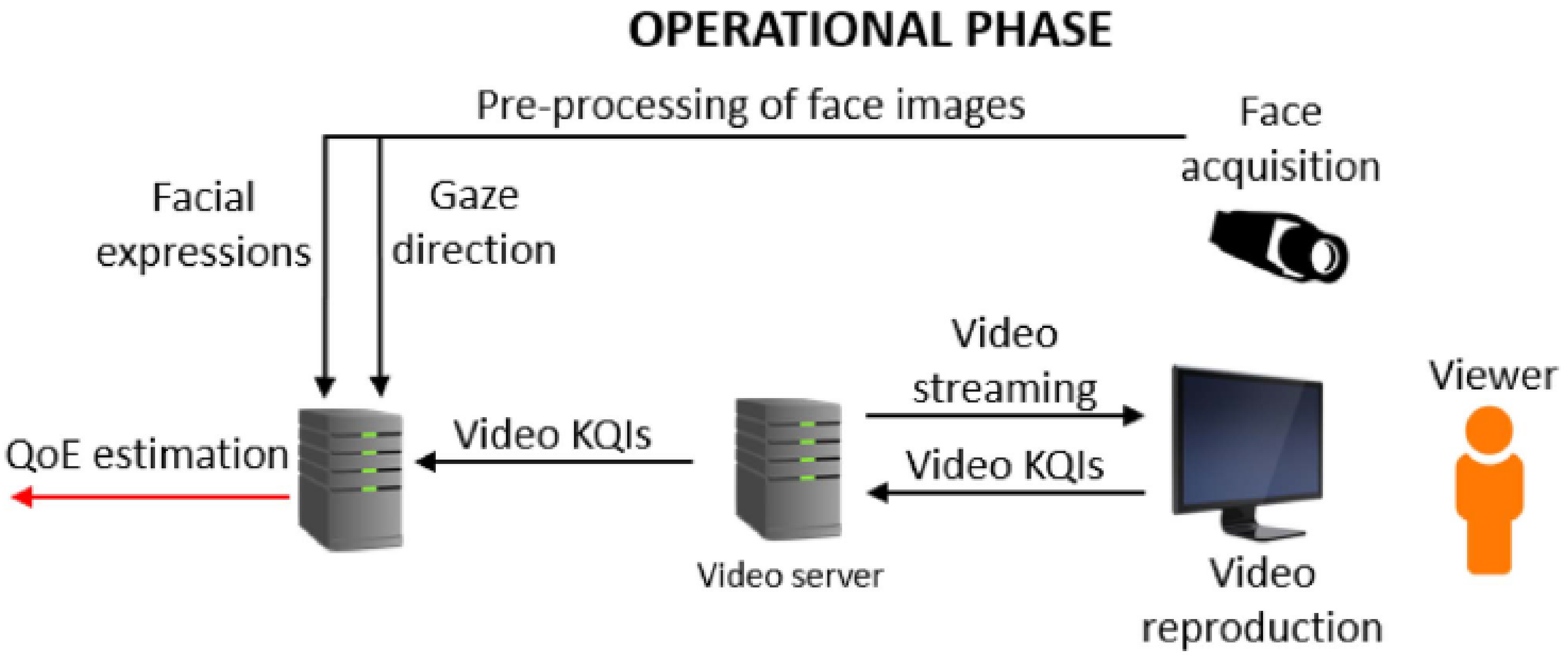
KQI: Key Quality Indicator
ACR: Absolute Category Rating

Video streaming session

- 1 - Crowdsourcing test
  - 20 neutral videos affected by long initial delays and re-buffering events.
- 2 - Laboratory test
  - 105 videos subject to impairment caused by blurring

S. Porcu, A. Floris, J. -N. Voigt-Antons, L. Atzori and S. Möller, "Estimation of the Quality of Experience During Video Streaming From Facial Expression and Gaze Direction," in *IEEE Transactions on Network and Service Management*, Dec. 2020

Prediction of QoE in HAS video streaming sessions



Automatically and unobtrusively QoE estimation jointly from

- Facial Expression and gaze direction
- KQIs

1 - Crowdsourcing Test

- Amazon Mturk platform
- HTML5 and JavaScript
- Face recorded only during the video execution
- Test videos: from the LIVE Mobile Stall Video Database
- Test conditions
  - **Original (OR)**: 30-second version of the original video content without initial delay and buffering interruptions
  - **Long Initial (LI)**: long initial delay that lasted randomly in the range 8 - 20 s
  - **Long Initial + Few Long Buffering (LIFL)**: long initial delay plus few (between 1 and 3) long (between 10 and 15 s) buffering events
  - **Long Initial + Many Short Buffering (LIMS)**: long initial delay plus many (between 4 and 7) short (between 2 and 4 s) buffering events

2 - Laboratory Test

- HTML5 and JavaScript
- Recorded only during the video execution
- Test conditions:
  - **BLR0:** original video without blurring impairment
  - **BLR5H and BLR5E**: video post-processed with a Gaussian blurring kernel with the size of 5×5 px and standard deviation (SD) of 3 covering respectively the second half of the video and the entire video
  - **BLR10H and BLR10E**: video post-processed with a Gaussian blurring kernel with the size of 10 × 10 px and SD of 3 covering respectively the second half of the video and the entire video
  - **BLR15H and BLR15E**: video post-processed with a Gaussian blurring kernel with the size of 15×15 px and SD of 3 covering the second half of the video and the entire video, respectively

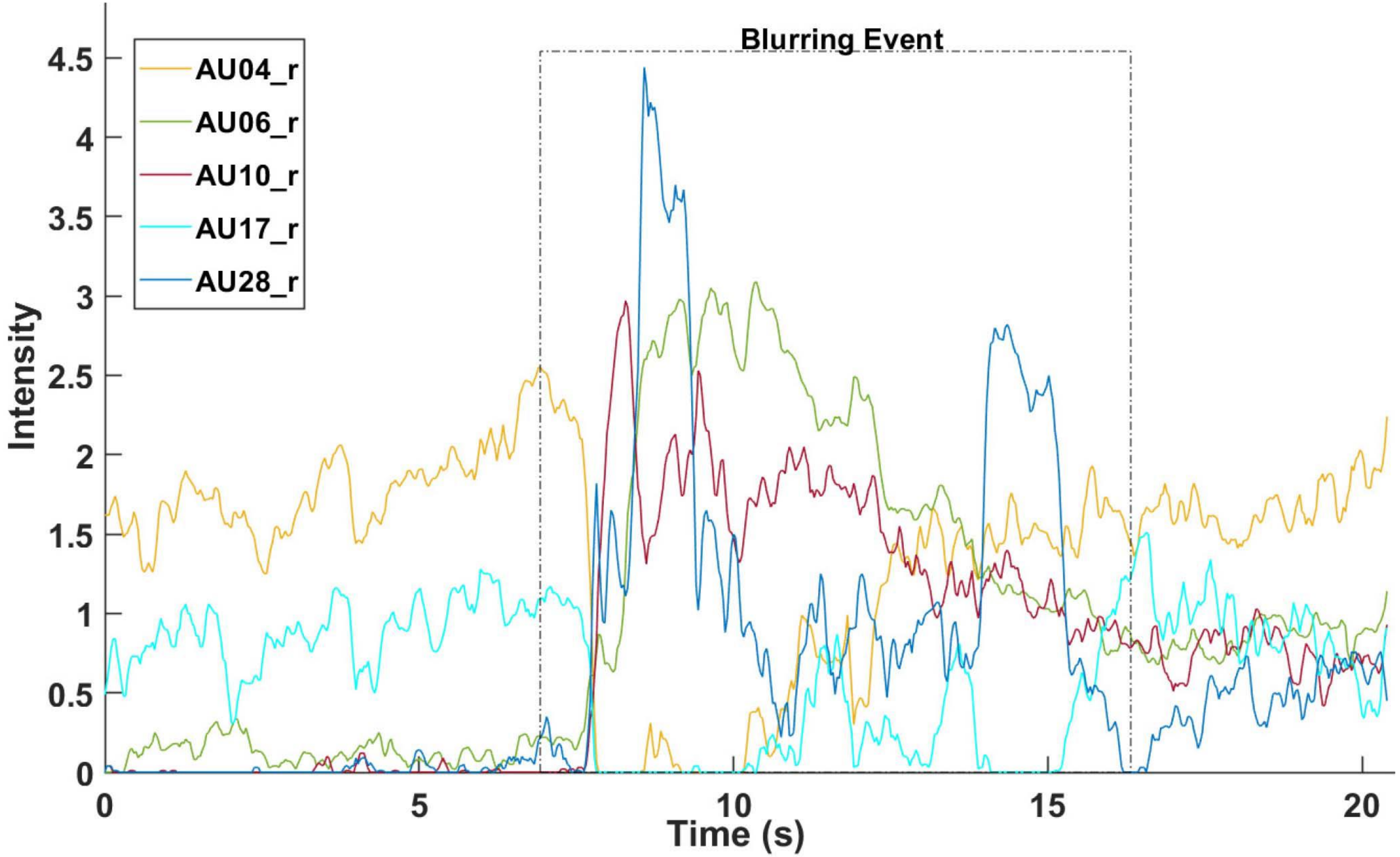
Data Preprocessing

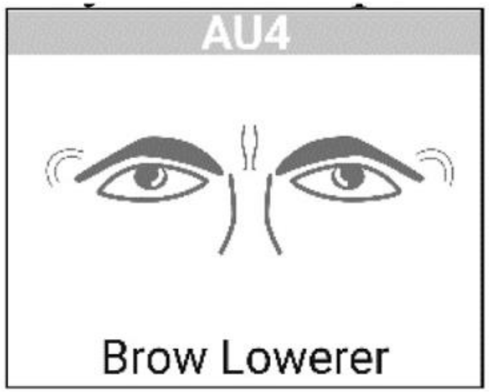
A. 3 metrics to make facial expressions and gaze direction independent from the duration of the recorded user's face videos
B. A devised impairment level feature that could put together different types of impairments when devising the previctor
   - Blurring
   - Delay and re-buffering

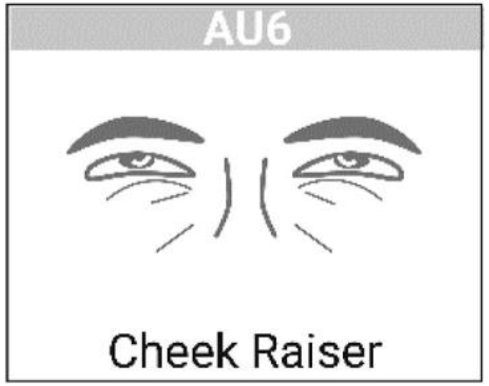
A - Data Preprocessing

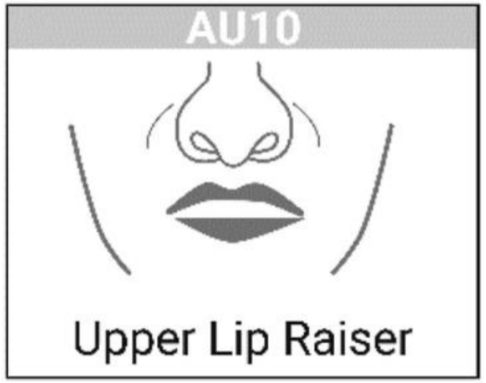
- AUs and Gaze direction extracted though OpenFace software (6 GD + 35 AUs)
- We computed 3 metrics:
  - Frequency of AU activation $F_{AU_c^j} = \frac{\sum_{n=1}^{N} a_n^j}{\sum_{n=1}^{N} \sum_{j=1}^{J} a_n^j}$
  - Intensity of the activated AUs $I_{AU_r^k} = \sum_{n=1}^{N} AU_{r,n}^k / N$
  - Variance of the Gaze Direction $V_{GD^g} = \frac{\sum_{n=1}^{N} (GD_n^g - \overline{GD^g})}{N}$
- We selected the AUs and Gaze features which were significant using ANOVA (p-value < 0.001): 6 GD and 25 AU finally used
- Some key AUs:
  - AU04 Brow Lowerer, AU06 Cheek Raiser, AU10 Upper Lip Raiser

A - Features temporal evolution

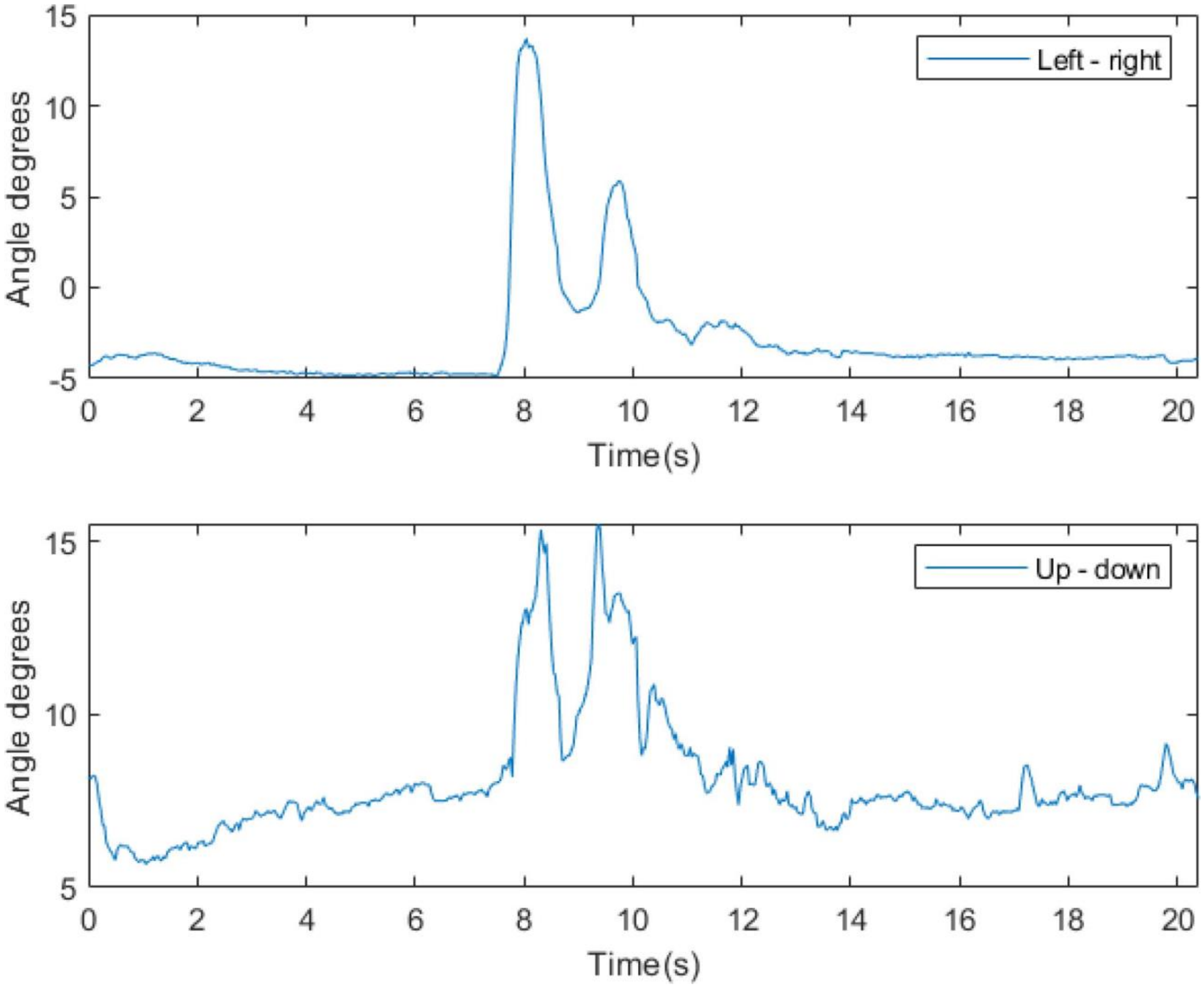




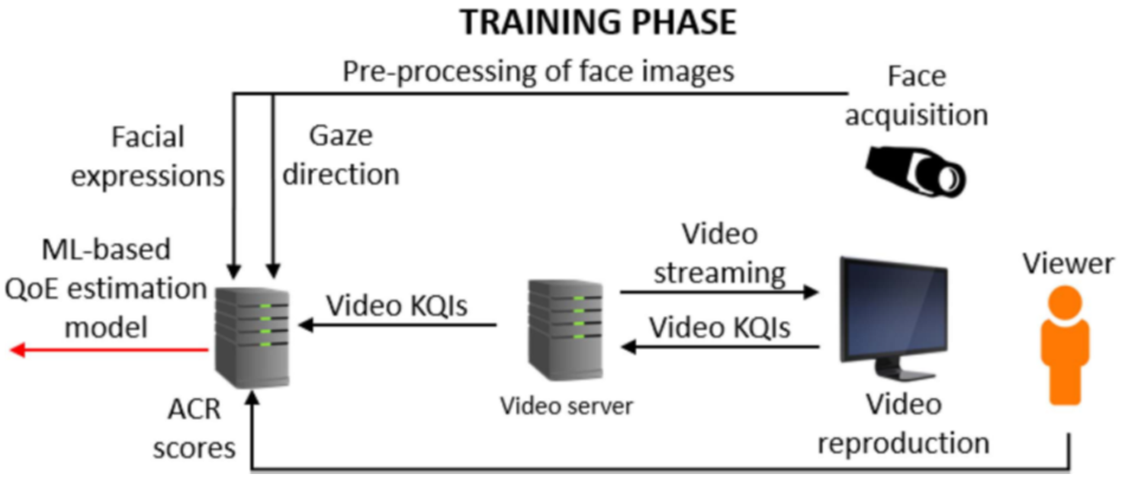




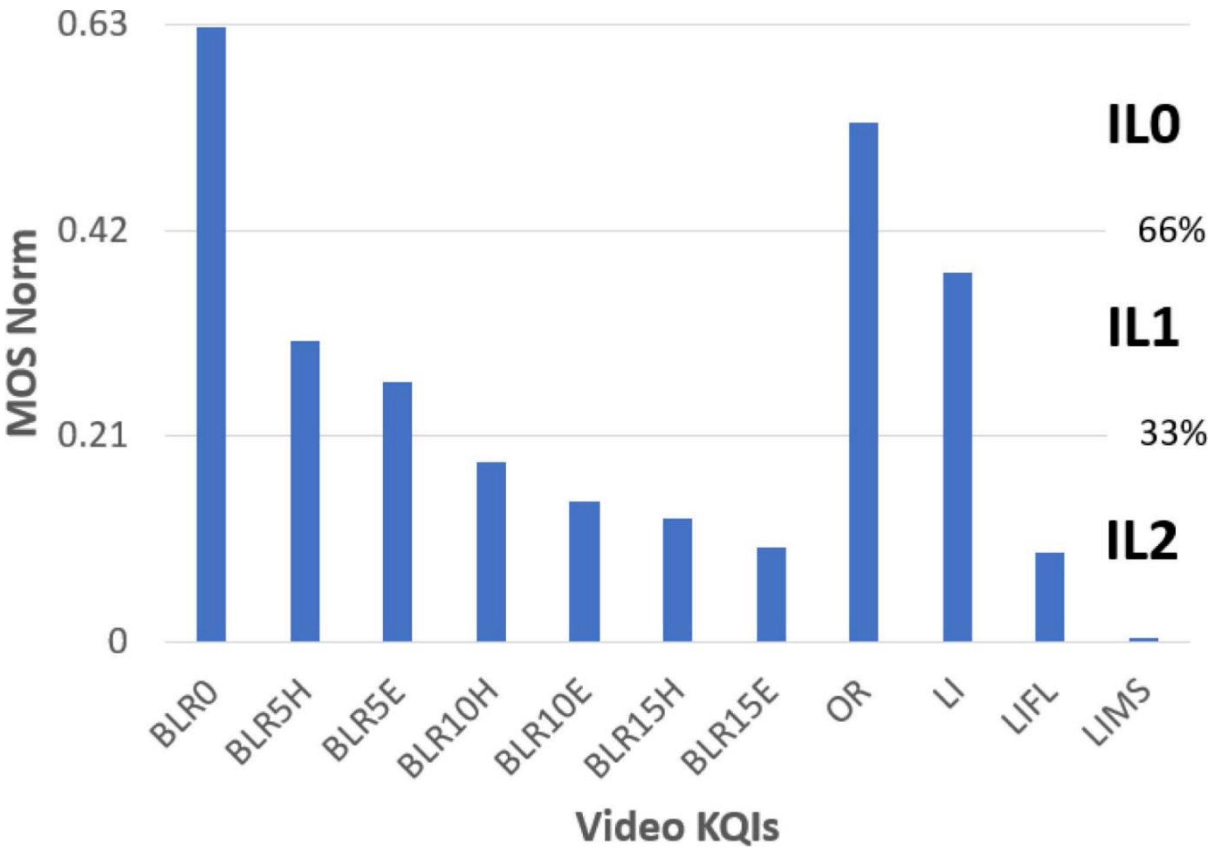
A - Features temporal evolution



B - Data Preprocessing



- Two different tests
  - Different impairments
- Our objective:
  - Develop a unified model
- We need a way to merge the dataset
  - Devise a common impairment index



Machine Learning Prediction Results

Model	Training dataset	Validation dataset	5-level quality scale			3-level quality scale			2-level quality scale		
			Specificity	Sensitivity	Accuracy	Specificity	Sensitivity	Accuracy	Specificity	Sensitivity	Accuracy
AU&GDtoQoE	Crowd augmented	Lab	1 = 98.00% 2 = 96.39% 3 = 88.43% 4 = 95.41% 5 = 92.77%	1 = 74.19% 2 = 79.17% 3 = 82.61% 4 = 83.33% 5 = 76.74%	78.0%	1/2 = 95.72% 3 = 93.20% 4/5 = 85.77%	1/2 = 80.56% 3 = 82.16% 4/5 = 87.70%	84.4%	1/2 = 95.36% 3/4/5 = 95.36%	1/2 = 91.39% 3/4/5 = 91.39%	91.4%
	Lab augmented	Crowd	1 = 88.44% 2 = 89.19% 3 = 96.94% 4 = 99.22% 5 = 99.90%	1 = 87.71% 2 = 83.01% 3 = 81.77% 4 = 74.10% 5 = 77.78%	83.5%	1/2 = 71.50% 3 = 98.76% 4/5 = 99.53%	1/2 = 98.45% 3 = 94.07% 4/5 = 89.82%	93.7%	1/2 = 91.17% 3/4/5 = 91.17%	1/2 = 97.84% 3/4/5 = 97.84%	95.6%
	70% Crowd+Lab augmented $k = 5$	30% Crowd+Lab augmented $k = 5$	1 = 91.43% 2 = 96.20% 3 = 97.69% 4 = 98.44% 5 = 98.69%	1 = 95.98% 2 = 86.56% 3 = 80.18% 4 = 79.40% 5 = 78.91%	87.8%	1/2 = 73.33% 3 = 98.50% 4/5 = 99.41%	1/2 = 98.15% 3 = 93.08% 4/5 = 88.86%	93.6%	1/2 = 94.85% 3/4/5 = 94.85%	1/2 = 99.02% 3/4/5 = 99.02%	96.8%
AU&GD&KQItoQoE	Crowd augmented	Lab	1 = 99.99% 2 = 98.92% 3 = 87.17% 4 = 95.11% 5 = 98.49%	1 = 88.57% 2 = 77.60% 3 = 93.75% 4 = 87.62% 5 = 77.23%	85.2%	1/2 = 97.63% 3 = 95.25% 4/5 = 91.77%	1/2 = 87.23% 3 = 89.78% 4/5 = 92.54%	90.5%	1/2 = 98.13% 3/4/5 = 98.13%	1/2 = 88.09% 3/4/5 = 88.09%	95.9%
	Lab augmented	Crowd	1 = 88.02% 2 = 91.50% 3 = 98.09% 4 = 99.52% 5 = 99.81%	1 = 91.76% 2 = 88.03% 3 = 79.80% 4 = 81.05% 5 = 87.50%	86.5%	1/2 = 98.41% 3 = 97.75% 4/5 = 99.99%	1/2 = 98.64% 3 = 97.78% 4/5 = 93.33%	97.7%	1/2 = 97.18% 3/4/5 = 97.18%	1/2 = 99.00% 3/4/5 = 99.00%	99.0%
	70% Crowd+Lab augmented $k = 5$	30% Crowd+Lab augmented $k = 5$	1 = 95.85% 2 = 96.18% 3 = 98.41% 4 = 98.61% 5 = 99.29%	1 = 96.87% 2 = 93.23% 3 = 90.26% 4 = 89.47% 5 = 89.67%	93.9%	1/2 = 92.17% 3 = 97.89% 4/5 = 99.35%	1/2 = 98.38% 3 = 96.08% 4/5 = 94.02%	97.1%	1/2 = 91.33% 3/4/5 = 91.33%	1/2 = 99.36% 3/4/5 = 99.36%	98.1%

Research questions:

1) *is there any correlation between facial expression and perceived Quality of Experience?*

Yes, ANOVA told us that there are 25 relevant AUs

2) can we predict QoE directly from facial expression?

Yes, with an accuracy of 93.9 when combined with KQI

3) can the predictor be application independent?

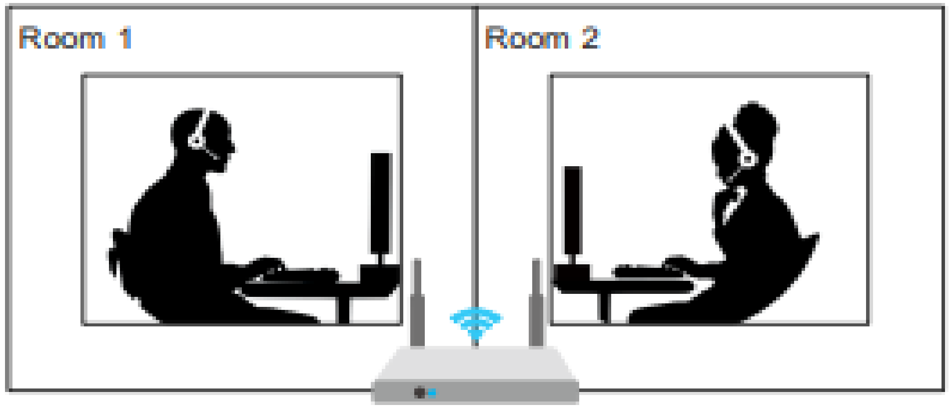
Only impairment independent … still to be studied

Application to the real-time video-call scenario

- The previous approach demonstrated it is possible to understand the perceived video quality of a video streaming session
- Is it possible to apply the same methodology to the real time video call scenario?
  - Video impairments are the same, but the user can behave differently due to the type of interactivity
    - Real-time interaction with other people
    - Not a passive video-watching session

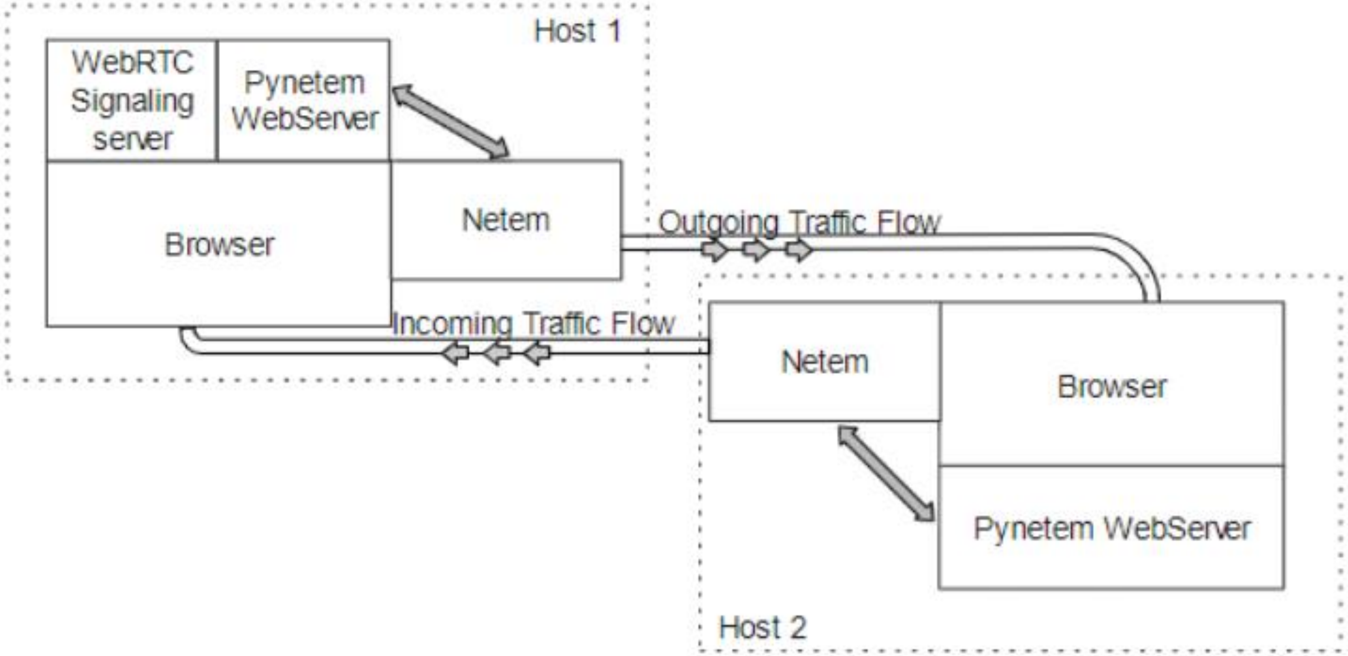
Application to WebRTC sessions

WebRTC considered scenario and test system architecture



The scenario of the experiment with the participants in different rooms

The two subjects took part to the celebrity name-guessing game



System architecture

Subjective quality assessment

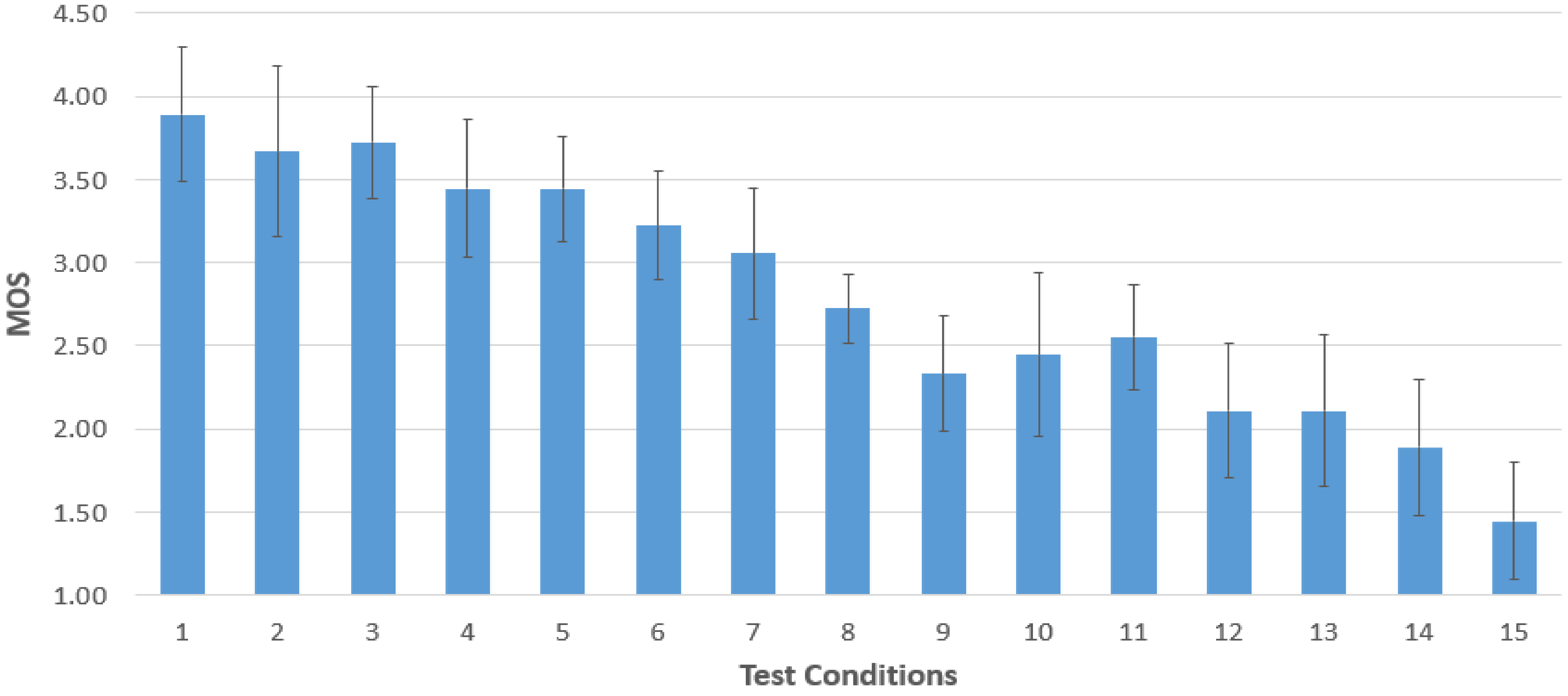
TC	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
Delay (ms)	0	500	1000	500	1000	0	500	1000	500	1000	0	500	1000	500	1000
Jitter (ms)	0	0	0	500	500	0	0	0	500	500	0	0	0	500	500
PLR (%)	0	0	0	0	0	15	15	15	15	15	30	30	30	30	30

a) *Network Impairments:*

- Delay: 0-500-1000 [ms]
- Jitter: 0-500 [ms]
- Packet Loss: 0-15-30 [%]

20 participants took part to the subjective tests

TC	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
Delay (ms)	0	500	1000	500	1000	0	500	1000	500	1000	0	500	1000	500	1000
Jitter (ms)	0	0	0	500	500	0	0	0	500	500	0	0	0	500	500
PLR (%)	0	0	0	0	0	15	15	15	15	15	30	30	30	30	30



Mean Opinion Score (MOS) with 95% confidence interval (CI)

G. Bingol et al., "The Impact of Network Impairments on the QoE of WebRTC applications: A Subjective study," 2022 14th International Conference on Quality of Multimedia Experience (QoMEX), 2022, pp. 1-6, doi: 10.1109/QoMEX55416.2022.9900882.

QoE Prediction Models

NUMBER OF ACR SCORES BEFORE AND AFTER DATA AUGMENTATION

ACR	Collected samples	Augmented samples
1	37	105
2	65	121
3	103	103
4	44	101
5	21	99
Total	270	529

Preprocessing of facial features:

- Data augmentation
- The adaptive synthetic **(ADASYN)** algorithm to achieve class over-sampling

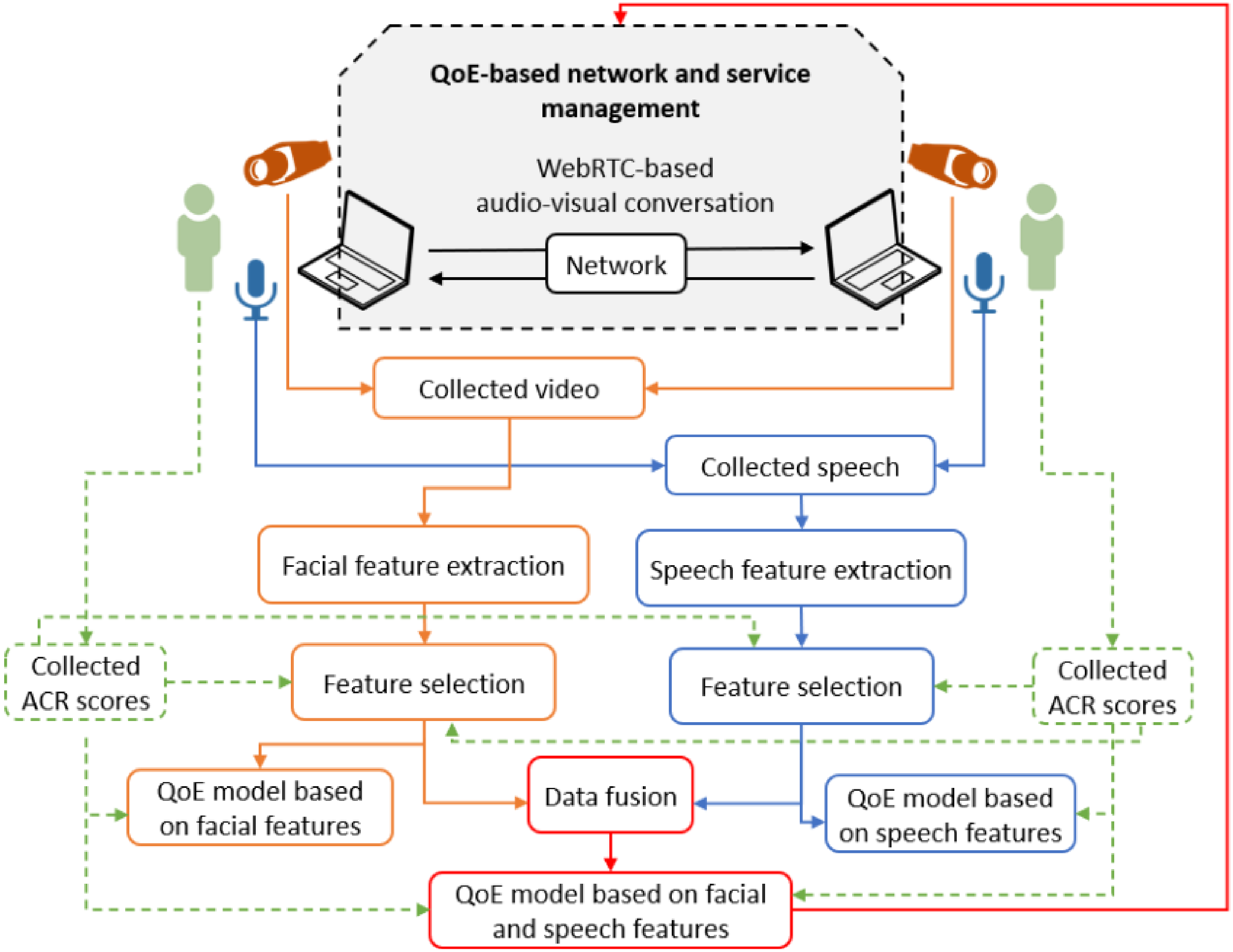
Performance of QoE Estimation Models

Estimation Model	Metric	ACR score				
		1	2	3	4	5
Method 1 k-NN	Mean Acc.			0.70		
	Accuracy	0.80	0.78	0.30	0.75	0.86
	Precision	0.71	0.72	0.49	0.65	0.86
	Recall	0.80	0.78	0.30	0.75	0.86
	F1-Score	0.75	0.75	0.37	0.70	0.86
Method 2 k-NN	Mean Acc.			0.60		
	Accuracy	0.72	0.57	0.20	0.65	0.88
	Precision	0.65	0.56	0.30	0.60	0.8
	Recall	0.72	0.57	0.20	0.65	0.88
	F1-Score	0.68	0.57	0.24	0.62	0.84
Method 3 Weighted k-NN	Mean Acc.			0.42		
	Accuracy	0.47	0.33	0.16	0.38	0.77
	Precision	0.45	0.36	0.17	0.38	0.67
	Recall	0.47	0.33	0.16	0.38	0.77
	F1-Score	0.46	0.34	0.17	0.38	0.72
Method 4 SVM	Mean Acc.			0.78		
	Accuracy	0.97	0.79	0.35	0.81	0.98
	Precision	0.86	0.71	0.62	0.75	0.89
	Recall	0.97	0.79	0.35	0.81	0.98
	F1-Score	0.91	0.75	0.45	0.78	0.93

- The best performance obtained with the SVM with an accuracy of 0.78
- Better results reached with the video on-demand scenario where an accuracy of 93.9 has been reached
  - More extended dataset
  - Difficulty in detecting the impairment

Estimation from (ALSO) the voice signal

Methodology



Speech files

- We considered 3 versions of the speech files:
  - **Original Speech (OS)**:
  - **Noise Reduced Speech (NRS)**: non-stationary noise reduction method (called spectral gating) to the original speech file to reduce the background noise
  - **Non-Silent Speech (NSS)**: removal of the silent intervals from the OS
- The **OpenSMILE** toolkit was used to extract speech features from the speech files
  - 64 low-level descriptor (LLD) specifically related to
    - energy characteristics (4)
    - spectral characteristics (55)
    - voicing characteristics (6)
- A total of 6373 functional statistical features are computed for the considered LLDs

Statistical analysis of speech features

- **One-way ANOVA** computed between the speech features and the corresponding ACR scores
- **Significance level p-value < 0.01**

LLD	OS	NRS	NSS
audspec_lengthL1norm	7	4	1
audSpec_Rfilt	8	49	10
audspecRasta	-	7	4
F0final	-	2	3
jitterLocal	1	-	9
jitterDDP	-	-	3
logHNR	-	3	2
mfcc_sma	43	20	25
pcm_fftMag	44	16	14
pcm_RMSenergy	3	3	3
pcm_zcr	5	3	-
shimmerLocal	-	-	4
voicingFinalUnclipped	2	4	8
Total	**113**	**111**	**86**

Data augmentation on speech features

- Data augmentation using ADASYN was performed to correct the dataset's class imbalance
- The number of augmented samples differs for the 3 speech files because it also depends on the different number of significant features found for each speech file

ACR score	Collected samples	Augmented Samples OS	NRS	NSS
1	37	104	102	107
2	65	107	110	107
3	103	103	103	103
4	44	111	104	105
5	21	104	98	97
Total	**270**	**529**	**517**	**519**

ML model based on speech features

Speech file / ML model	Performance metric	ACR score 1	2	3	4	5
OS SVM	Mean Acc.			0.83		
	Precision	0.85	0.71	0.76	0.80	0.99
	Recall	0.95	0.76	0.52	0.89	0.99
	F1-Score	0.90	0.73	0.62	0.85	0.99
NRS SVM	Mean Acc.			0.86		
	Precision	0.98	0.89	0.61	0.99	0.99
	Recall	0.93	0.70	0.88	0.83	0.99
	F1-Score	0.95	0.78	0.72	0.90	0.99
NSS SVM	Mean Acc.			0.85		
	Precision	0.93	0.79	0.76	0.83	0.93
	Recall	0.93	0.85	0.57	0.91	0.98
	F1-Score	0.93	0.82	0.65	0.87	0.95

When using speech features

Recall that when using facial expression and gaze direction we reached an accuracy as high as <u>0.78</u> with SVM

ML model based on facial and speech features – Data fusion

- The SVM is the ML classifier that demonstrated to achieve the best QoE estimation performance on the individual facial and speech features datasets
- We utilized an SVM as the ML classifier and we considered two data fusion approaches to fuse the $\boldsymbol{FAC_{aug}}$ and $\boldsymbol{SP_{aug}}$ datasets:
  - **Principal Component Analysis (PCA)**: statistics technique used to transform the original dataset into a reduced dataset of new variables capturing the most important patterns and relationships in the data
  - **Improved Centered Kernel Alignment (ICKA)**: method used for ML feature fusion tasks, which computes the SVM kernel alignment between the ideal kernel blocks and the base kernel that are selected to be representative of the data. The fusion kernel $\boldsymbol{ICKA_{kernel}}$ is built as a weighted linear combination of multiple aligned kernels, and it is used as the input kernel of the SVM

ML model based on facial and speech features – performance

- QoE estimation performance achieved by the SVM model trained with facial and speech features fused with PCA and ICKA techniques

Data fusion technique	Performance metric	ACR score 1	2	3	4	5
PCA	Mean Acc.			0.84		
	Precision	0.95	0.74	0.60	0.93	0.95
	Recall	0.90	0.85	0.71	0.86	0.90
	F1-score	0.93	0.78	0.68	0.87	0.93
ICKA	Mean Acc.			0.93		
	Precision	0.85	0.83	0.82	0.95	0.87
	Recall	0.92	0.92	0.91	0.99	0.93
	F1-score	0.87	0.85	0.80	0.95	0.88

ML model based on facial and speech features – performance

- Comparison of QoE estimation performance achieved by ML models trained on facial features only, speech features only, and combined facial and speech features

ML model	Features	Mean accuracy
SVM	Facial	0.78
SVM	Speech (NRS)	0.86
Data fusion PCA	Facial + speech	0.84
Data fusion ICKA	Facial + speech	0.93

UNIVERSITY OF CAGLIARI

DIEE - Department of Electrical and Electronic Engineering

Are QoE and Sustainability reconcilable?

Net4U

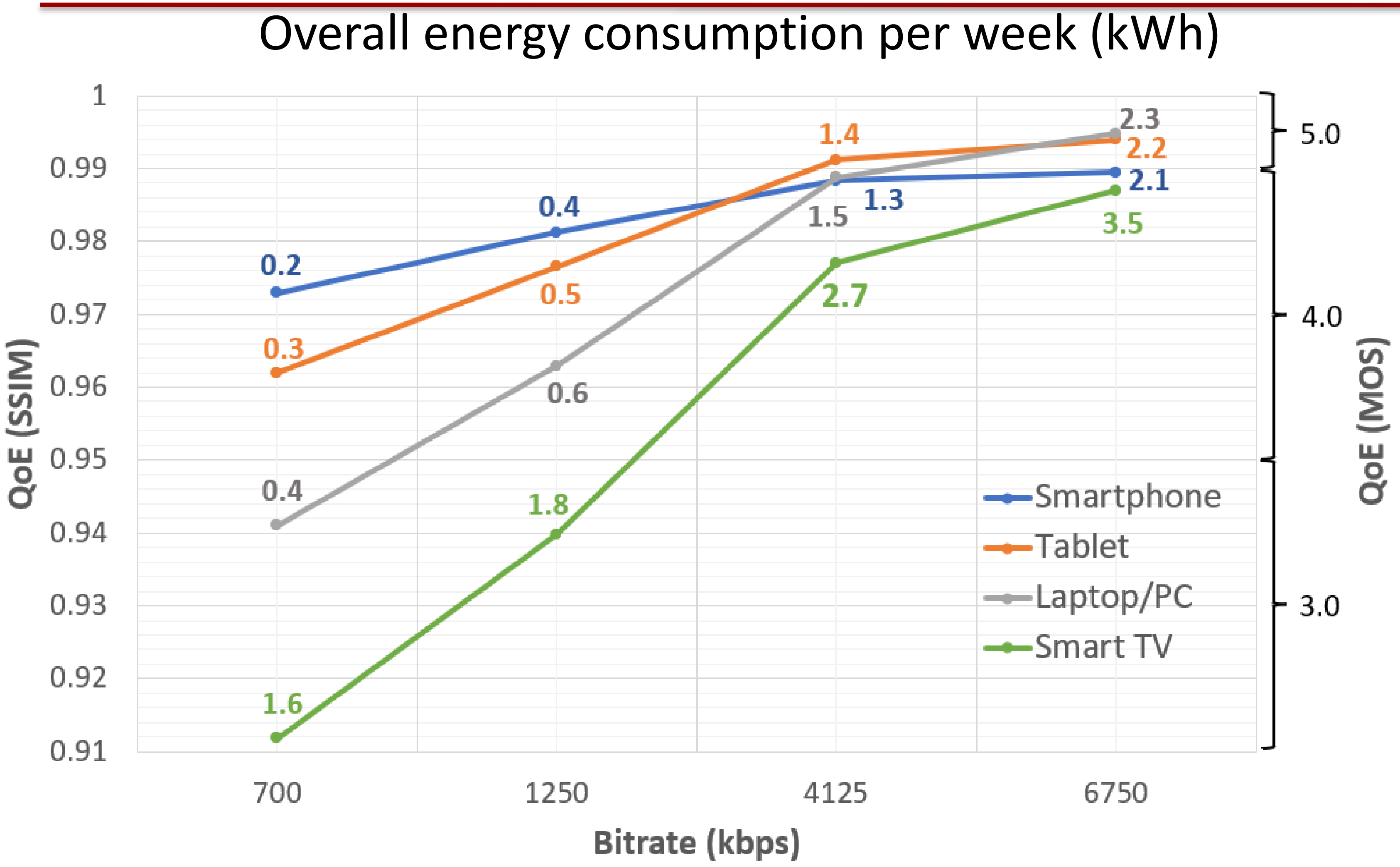
Networks for Humans laboratory: https://sites.unica.it/net4u

Status of energy consumption

- **ICT's current share of global greenhouse gas (GHG) emissions is estimated to be between 2% and 4%** [1]
  - 35%-60% end-devices, 25%-35% networks, 20%-40% data centers
- **Expansion in the delivery of video contents in recent years**
  - Spread of streaming services, social networks, higher-resolution content, different end-devices
  - Video traffic alone accounted for the 82% of all consumer Internet traffic in 2022, while it accounted for the 75% in 2017 [2]

[1] C. Freitag, M. Berners-Lee, K. Widdicks, B. Knowles, G.S. Blair, A. Friday, The real climate and transformative impact of ICT: A critique of estimates, trends, and regulations, Patterns, Vol. 2, Issue 9, 2021.
[2] T. Barnett, S. Jain, U. Andra, and T. Khurana, "Cisco visual networking index (vni) complete forecast update, 2017–2022," 2018.

Bitrate vs quality vs energy consumption



- Standard resolution per device
- QoE as a function of the bitrate
- No network impairments
- 7-hour streaming per day

Deeper investigation

Research question

What are the factors that impact on energy consumption and user's QoE during video streaming?

Is there a trade-off between acceptable QoE and green choices for users?

G. Bingöl, A. Floris, S. Porcu, C. Timmerer and L. Atzori, "Are Quality and Sustainability Reconcilable? A Subjective Study on Video QoE, Luminance and Resolution," *2023 15th International Conference on Quality of Multimedia Experience (QoMEX)*

Methodology

We designed and conducted a subjective assessment to investigate:

Different end devices *(TV, Laptop, and Smartphone)*

The impact of the different end devices on the QoE and energy consumption during video streaming

Different Video Resolution for each end device (4K, FHD and HD)

The impact of the video resolutions on the QoE and energy consumption during video streaming

Different types of Luminance Features *(Backlight, Ambient, and Content)*

The impact of different types of luminance features on the QoE and energy consumption during video streaming

Test conditions

Backlight luminance (BL)

BL = Min. *TV*: 300 lx, *LP*: 400 lx, *SP*: 200 lx

BL = Max. *TV*: 5500 lx, *LP*: 4000 lx, *SP*: 5000 lx

Ambient luminance (AL)

dark environment: 0 lx;

bright environment: 500 lx

Content luminance (CL)

Earth Mover's Distance (EMD) metric:

Capture the luminance over the frames

TC	BL (lx)			AL (lx)	RES			CL
	TV	LP	SP		TV	LP	SP	
TC1	300	400	200	0	4K	FHD	FHD	H
TC2					FHD	HD	HD	
TC3					4K	FHD	FHD	L
TC4					FHD	HD	HD	
TC5				500	4K	FHD	FHD	H
TC6					FHD	HD	HD	
TC7					4K	FHD	FHD	L
TC8					FHD	HD	HD	
TC9	5500	4000	5000	0	4K	FHD	FHD	H
TC10					FHD	HD	HD	
TC11					4K	FHD	FHD	L
TC12					FHD	HD	HD	
TC13				500	4K	FHD	FHD	H
TC14					FHD	HD	HD	
TC15					4K	FHD	FHD	L
TC16					FHD	HD	HD	

The Impact of the Electricity Consumption

- Device energy consumption
  - video resolution does not significantly impact the power load of the device, except for the SP
  - The BL has a significant impact on the power load of the device, which increases with the device's screen size
- Total energy
  - Video resolution has a significant impact, the biggest one on the TV
  - Watching a 4K video instead of an FHD video on the TV results in overall electricity consumption increase by 3.4 (low BL) times and 2.4 times (high BL)

$$Q_i = t_i \cdot (P_i + \rho \cdot R_i)$$

Device power *Total energy*

Device	RES	R_i (GB/h)	Avg BR (Mbps)	BL (lx)	P_i (W)	Q_i (kWh)
TV	4K	9.00	20	300	55.78	6.69
				5500	236.05	7.95
	FHD	2.25	5	300	53.67	1.95
				5500	243.48	3.28
LP	FHD	2.25	5	400	13.04	1.67
				4000	18.56	1.70
	HD	0.90	2	400	13.21	0.72
				4000	18.53	0.76
SP	FHD	2.25	5	200	1.85	1.59
				5000	2.46	1.59
	HD	0.90	2	200	2.98	0.65
				5000	3.19	0.65

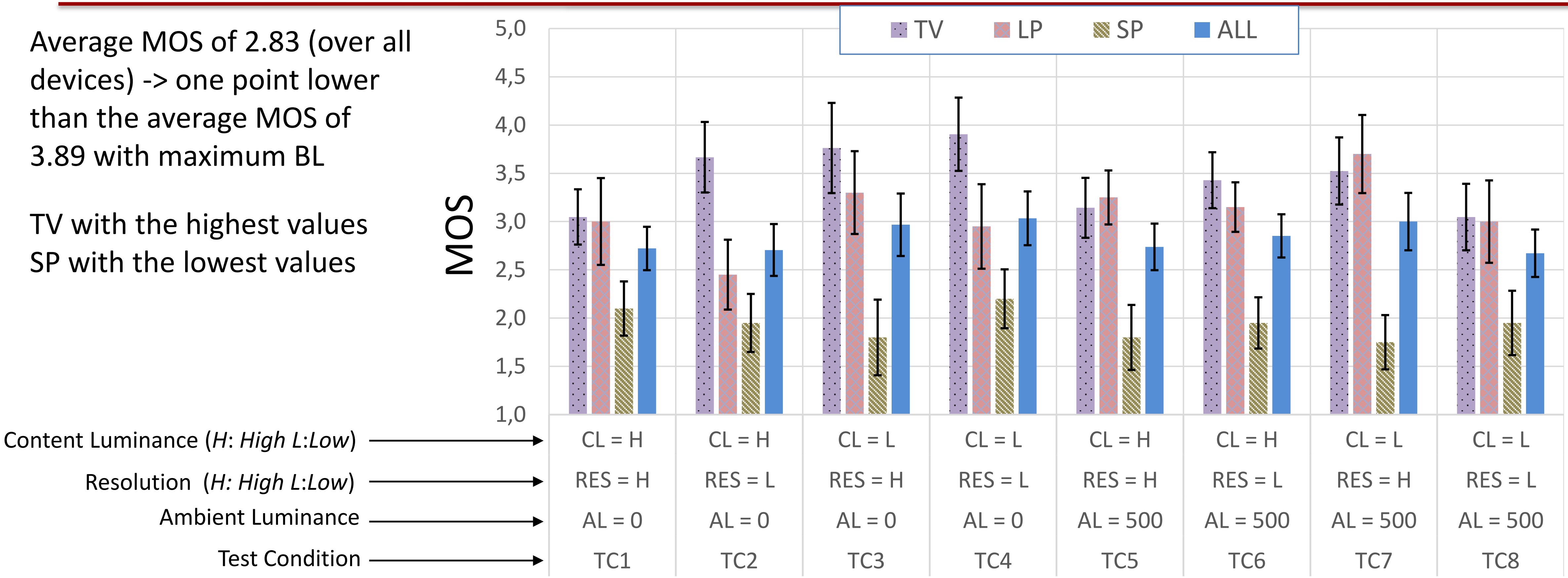
[1] P. Suski, J. Pohl, and V. Frick, "All you can stream: Investigating the role of user behavior for greenhouse gas intensity of video streaming," in Proc. of the 7th Int. Conf. on ICT for Sustainability, 2020, pp. 128–138.

[2] R. Madlener, S. Sheykhha, and W. Briglauer, "The electricity-and CO2-saving potentials offered by regulation of European video-streaming services," Energy Policy, vol. 161, p. 112716, 2022.

MOS Values with Minimum Backlight Luminance

Average MOS of 2.83 (over all devices) -> one point lower than the average MOS of 3.89 with maximum BL

TV with the highest values
SP with the lowest values



Resolution (*H: High*): ***TV***: *4K*, ***LP***&***SP***: *FHD*
Resolution (L: Low): ***TV***: *FHD*, ***LP***&***SP***: *HD*

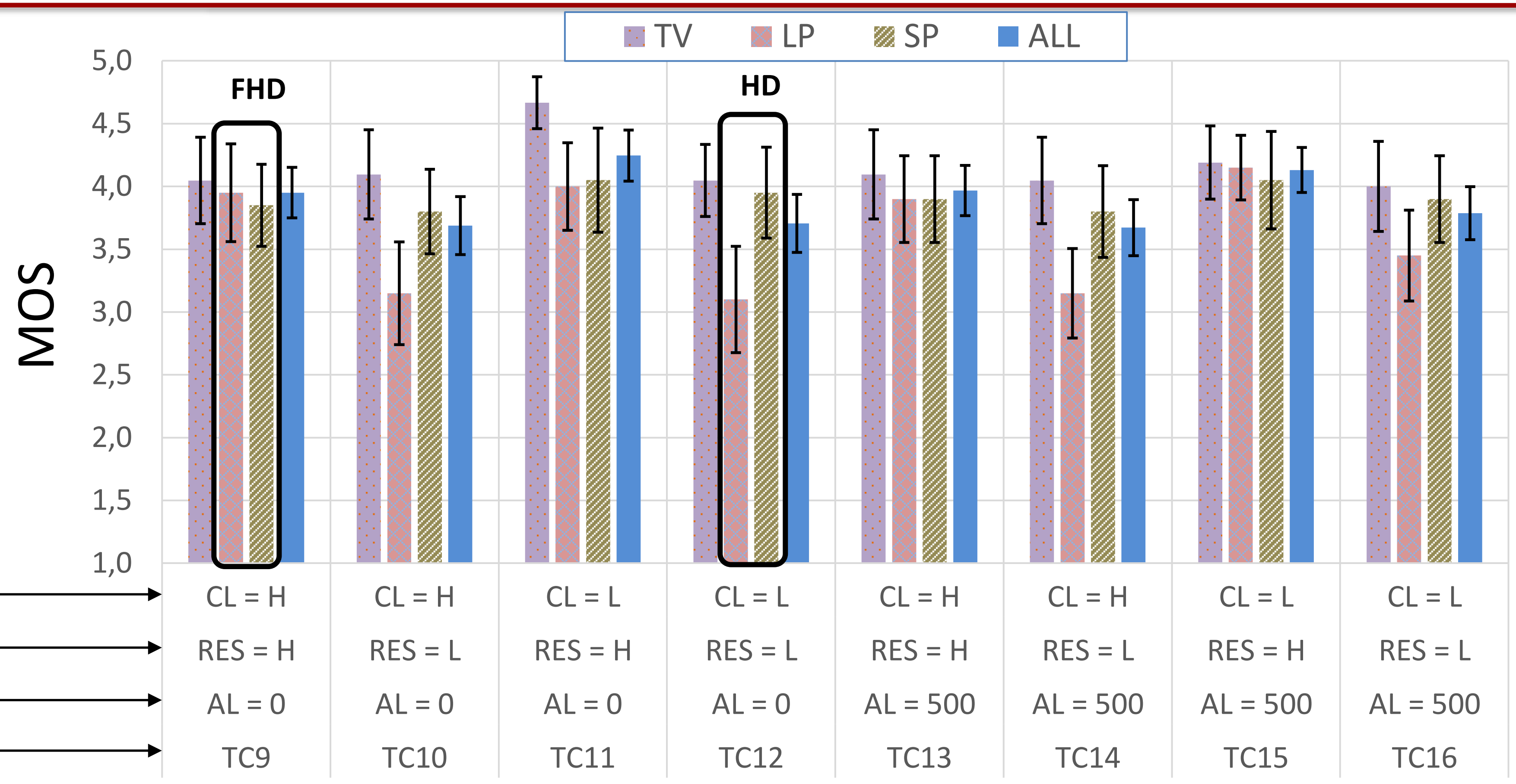
Mean Opinion Score (MOS) with 95% confidence interval (CI) for the **first** 8 TCs.

Figure: a) Devices' screens with Backlight = Min. TV: 300 lx, LP: 400 lx, SP: 200 lx

MOS Values with Maximum Backlight Luminance

SP: HD / FHD videos of comparable quality

On SP, low impact of low resolution



Content Luminance (*H*: *High L*:*Low*) →

Resolution (*H: High L*:*Low*) →

Ambient Luminance →

Test Condition →

Resolution (*H: High*): ***TV***: *4K*, ***LP&SP***: *FHD*

Resolution (L: Low): ***TV***: *FHD*, ***LP&SP***: *HD*

Mean Opinion Score (MOS) with 95% confidence interval (CI) for the **second** 8 TCs.

Figure 1: b) Devices' screens with Backlight = Max. TV: 5500 lx, LP: 4000 lx, SP: 5000 lx

Observations: QoE vs Sustainability

What's the impact of Backlight Luminance (BL) and Ambient Luminance (AL)?

- ✓ **setting the BL to the minimum** saves battery energy but decreases the user's QoE in any configuration

- ✓ **In a dark ambient (AL= 0)**
  - 4K videos: the BL significantly influences the QoE
  - FHD videos: the BL does not influence the QoE: the user may choose to set the BL to a minimum, saving up to **4 times** of the TV power load

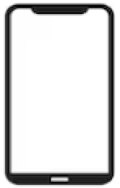
- ✓ **In a bright ambient (AL= 500)**
  - the BL significantly influences the QoE for all cases

- ✓ BL has not an impact when streaming HD videos with low CL
  - ✓ either with bright or dark ambient

Observations: QoE vs Sustainability

What’s the impact of Resolution (RES)?

- ✓ **Setting the RES to lower values**
  - – contributes to saving energy and CO2 emissions
  - – lower resolutions may negatively impact the user’s QoE

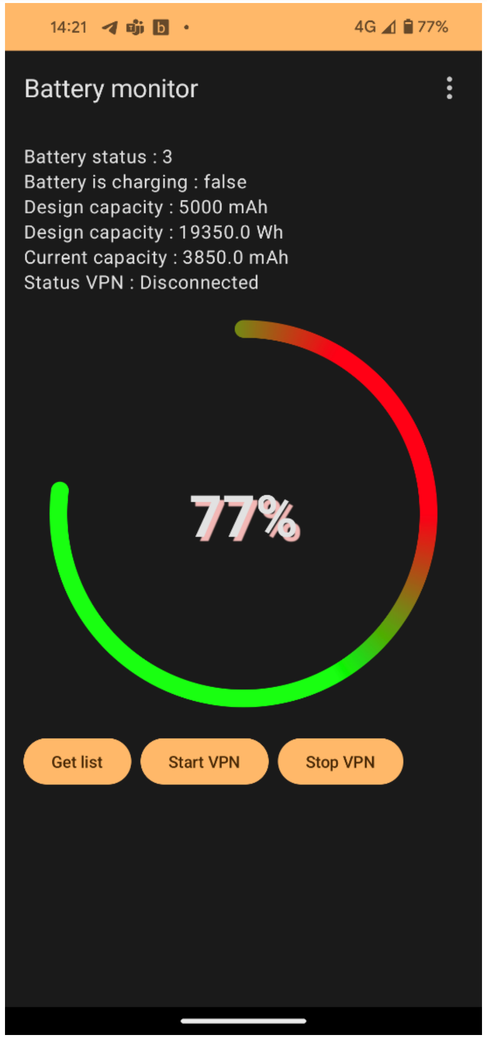


- ✓ No significant perceived QoE decrease with different video resolutions is observed
- ✓ **Watching HD rather than FHD content** ➡ can save energy and CO2 emissions

- ✓ **Significant differences for the perceived QoE is only found**
  - – bright content on a darker screen or dark content on a bright screen (optimal condition)

New campaign going on

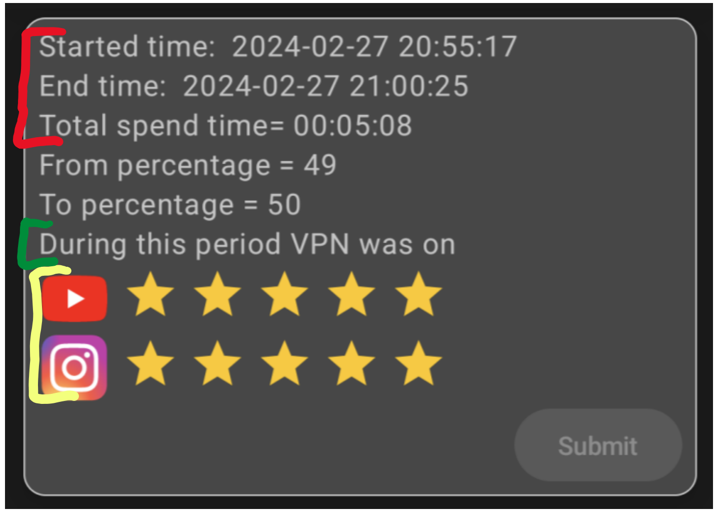
- Pursuing the goal of reducing the video streaming power consumption, we developed *Battery Monitor*
  - Android application for monitoring the smartphone consumptions and the network resource usage per application
  - It also asks feedback per video sessions
    - Youtube and TikTok
  - It allows for setting a limit to the network performance

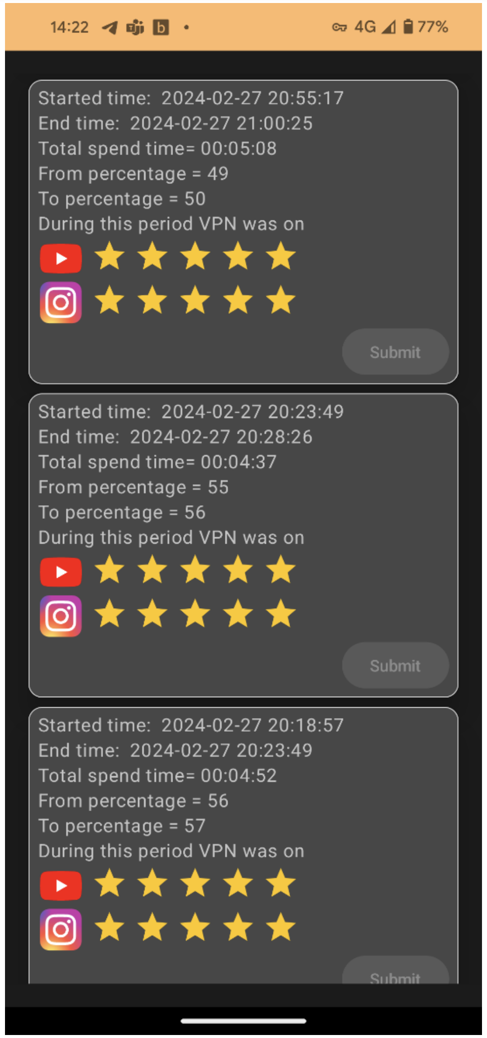


App features

- It allows to evaluate the QoE of multimedia video applications, giving you information regarding the period of usage of the application, how much data has been used for each session and if the network was limited or not.

- Statistics on video usage: time spent, resolution, bandwidth, energy consumption
- Perceived quality when reducing the bandwidth





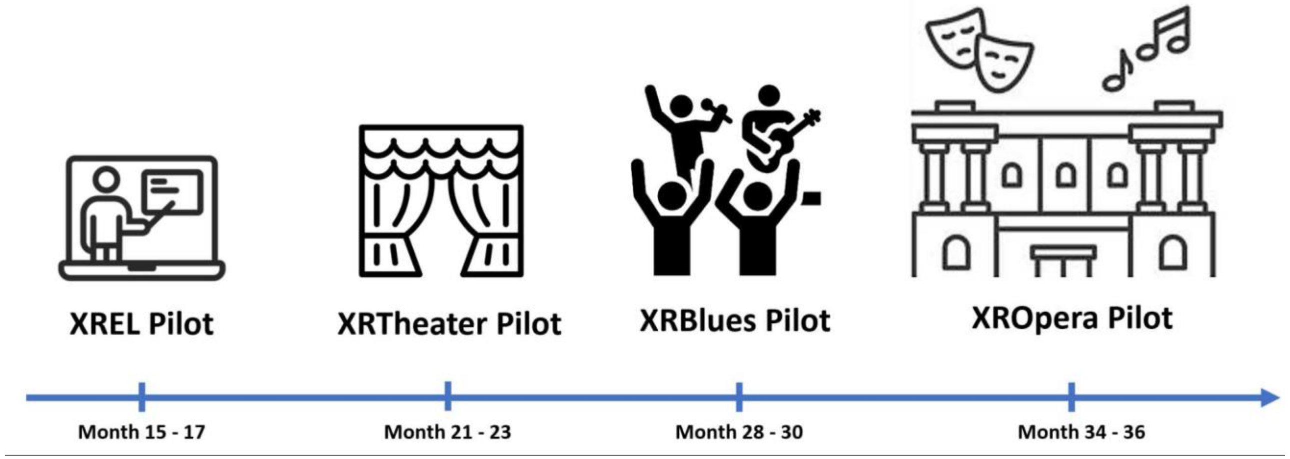
Conclusions

Takeoff messages

- Facial expression, voice analysis and gaze direction
  - High QoE estimation accuracy using predictors trained for specific appl.
  - Generalization has not been demonstrated yet
    - Tests with other application scenarios needed
    - Need a way to the combine influence factors -> Multi-view learning?
- Find an optimal trade-off between QoE and resources
  - To reach more sustainable multimedia services we need to involve the user
    - More tests are needed on the user behavior
    - Preliminary results show that a "green user" button can be effective
    - Gamification approaches needed

Ongoing activities

- Focus on XR applications
  - HE RIA, HEAT project, beg. June24
  - NextGenerationEU, HuTwin, March24



- Extract facial expression with the HMD
- QoE from user movements and behaviours
- Create the human digital twin

UNIVERSITY OF CAGLIARI
DIEE - Department of Electrical and Electronic Engineering

Thank you for your attention!

This activities have been carried out in collaboration with my group members, in particular: Gulnaz Bingol, Alessandro Floris, Simone Porcu

These works were partially supported by the European Union-NextGenerationEU (National Sustainable Mobility Center CN00000023, Italian Ministry of University and Research Decree n. 1033-17/06/2022, Spoke 8) and by the Italian Ministry of Enterprises and Made in Italy (MIMIT), within the 5G technology support program, on axis 1 "House of Emerging Technologies" (CTE), Project Name "Cagliari Digital Lab" (ID: G27F22000040008).