Influence of cross-modal IP-based degradations on the perceived audio-visual quality

Mylène C.Q. Farias, Andrew Hines[2] and Helard Martinez[1]
[1]University of Brasília, Brazil
[2]University College Dublin, Ireland
http://www.ene.unb.br/mylene

VQEG Meeting, Nov-11-2019, California

Universidade de Brasília

Contents

Quality of Experience (QoE) for Multimedia Content

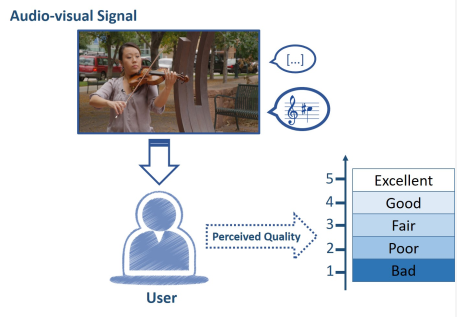
- Most MM content has audio and video!
- Audio and Video degradations.

Subjective Experiments

Main goals of this project:

- Design a NR pixel-based audio-visual quality metric;
- Study effect of both and audio degradations on audio-visual quality;
- Study cross-modal interactions;
- Create a large audio-visual dataset, with a diverse content and cross-modal degradations.

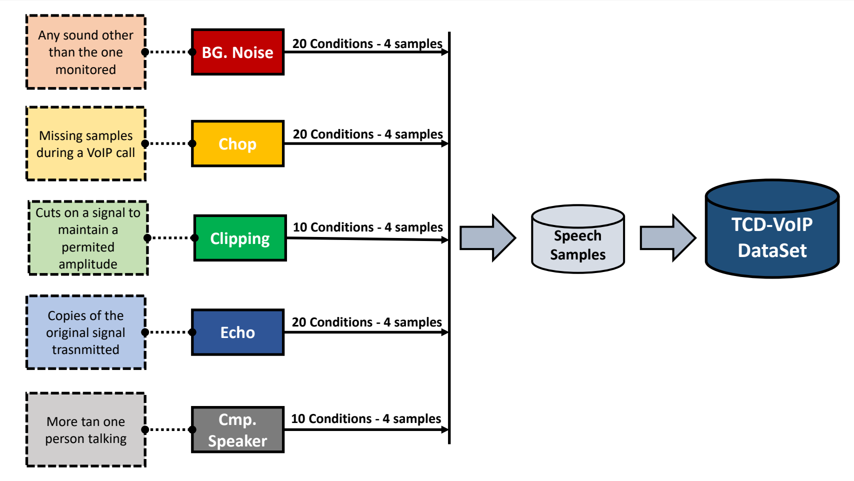
Subjective Experiments



- Experiment 1: Audio-visual signals with video degradations;
  - Video coding, Packet loss, Frame freezing
- Experiment 2: Audio-visual signals with audio degradations;
  - BG Noise, Chop, Clipping, Echo
- Experiment 3: Audio-visual signals with both audio and video degradations.

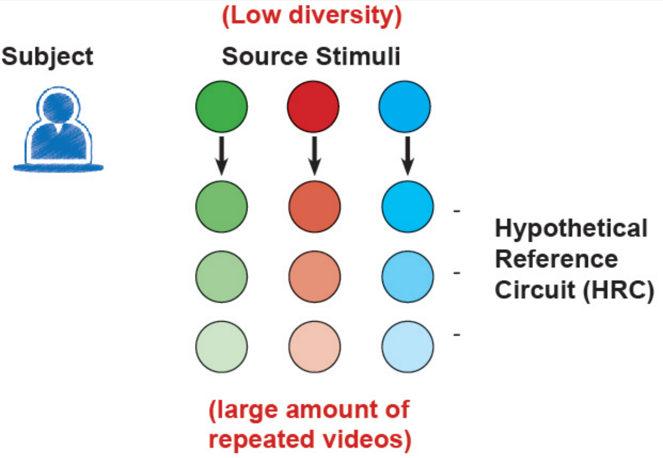
Contents

TCD-VoIP DataSet



- Audio dataset- Andrew Hines, University College Dublin
- Only four degradations were used (BG Noise, Chop, Clipping, and Echo).

Traditional Experimental Methodology

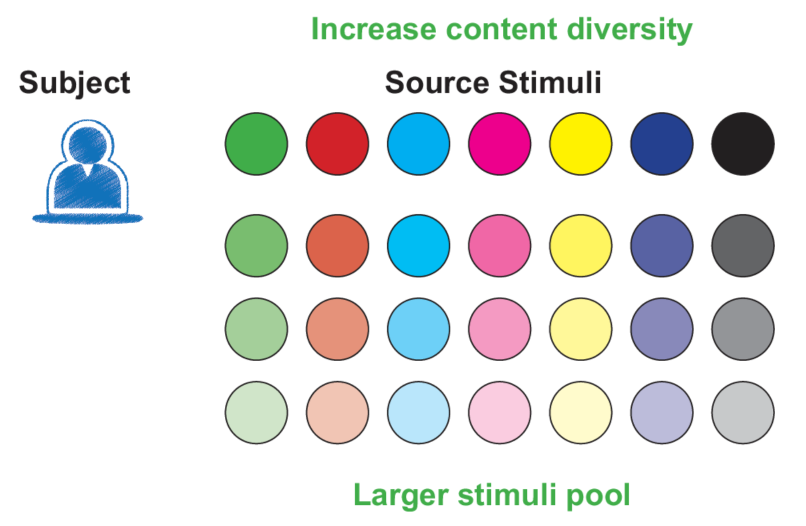


- Artificial Scenario
- Low content Diversity
- Short-length Sequences

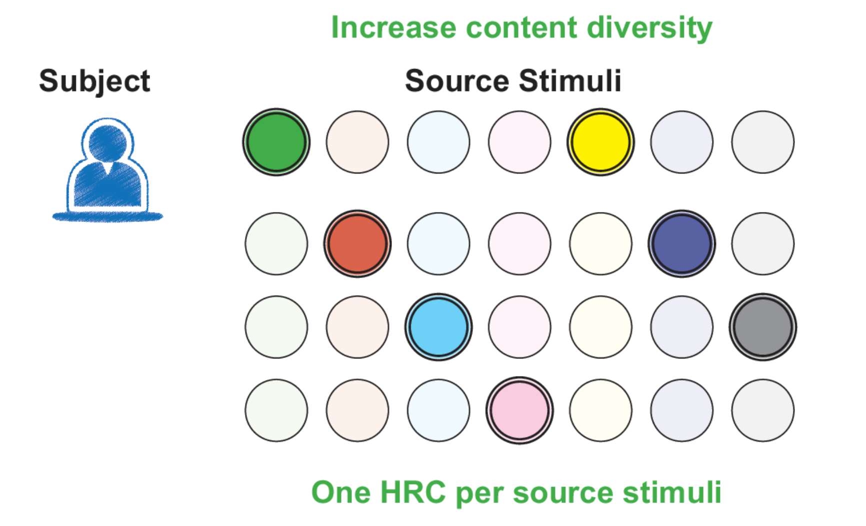
Immersive Methodology (IM) - M. Pinson

- Goals:
  - Increase content diversity;
  - Keeping the experiment interesting or/and more realistic;
  - Reduce fatigue.
- Longer stimuli (30 - 60 seconds):
  - Capture participant's attention;
  - Transmit an entire idea.
- Audio-visual stimuli:
  - Rate the global audio-visual quality;
  - Measure both quality and comfort.

Immersive Methodology (IM)



Immersive Methodology (IM)



Apparatus and Physical Conditions

- Experiment divided into 3 sessions: Display, Training, Main;
- Scores collected (ACR scale, 5 points): MQS_{HRC} - Mean Quality Score (HRC)
- Recording Studio @ University of Brasilia
- Desktop computer, LCD monitor, set of earphones, Sound card Asus Xonar DGX 5.1
- Viewing conditions: ITU Rec. BT.500
- Sixty (60) volunteers

Stimuli

Source stimuli:	40 HD sequences
Temporal resolution:	1280x720 (720p)
Spatial resolution:	30 fps
Color space format:	4:2:0
Average Length:	34 seconds
Bit-depth:	16 bits
Sample frequency:	48 kHz
Audio Codec:	PCM

Stimuli

- Video distortions: bitrate compression, Packet-Loss, and Frame-Freezing;
  - H.264 and H.265 video codecs (400 to 16,000 kbs);
  - Packet Loss (0.01 to 0.08)
  - Freezing Pauses (1, 3) and Length (2, 7)
- Four types of audio impairments: BG Noise, Chop, Clipping, Echo;
  - BG Noise (15, 10 dB)
  - Chop (rate 2 or 5 chop/s)
  - Clipping (multiplier by 11 or 25)
  - Echo (100 and 180 ms)

Table: Coding parameters and types of degradations of the video component of each HRC of the dataset.

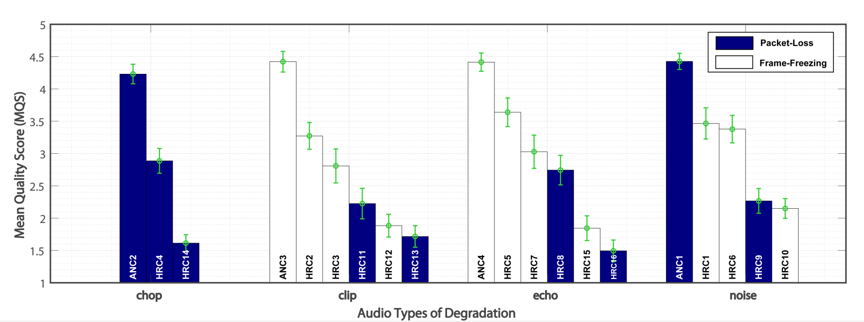
PacketLoss	Video Codec	Bitrate	PLR
HRC3	H.265	8000	0.01
HRC4	H.265	8000	0.01
HRC7	H.265	8000	0.01
HRC8	H.264	2000	0.05
HRC9	H.264	2000	0.05
HRC11	H.264	2000	0.05
HRC13	H.265	400	0.08
HRC14	H.265	400	0.08
HRC16	H.265	400	0.08
ANC1	-	-	-
ANC2	-	-	-
Frame Freezing	**Video Codec**	**Coding Bitrate**	**Freezing Pauses (P), Length (L)**
HRC1	H.264	16000	$P = 1, L = 2$
HRC2	H.264	16000	$P = 1, L = 2$
HRC5	H.264	16000	$P = 1, L = 2$
HRC6	H.264	16000	$P = 1, L = 2$
HRC10	H.264	800	$P = 3, L = 7$
HRC12	H.264	800	$P = 3, L = 7$
HRC15	H.264	800	$P = 3, L = 7$
ANC3	-	-	- -
ANC4	-	-	- -

Table: Coding parameters and types of degradations of the audio component for each HRC of the dataset.

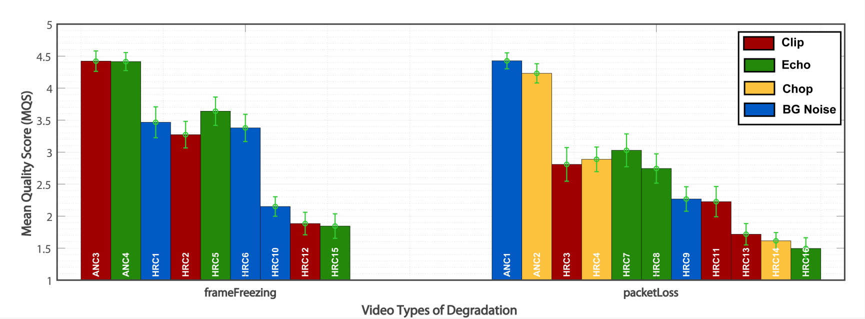
BG Noise	**Noise**	**SNR (dB)**	
HRC1	car	15	
HRC6	office	10	
HRC9	office	10	
HRC10	office	10	
ANC1	-	-	
Chop	**Period (s)**	**Rate (chop/s)**	**Mode**
HRC4	0.02	2	zeros
HRC14	0.02	5	zeros
ANC2	-	-	-
Clip		**Multiplier**	
HRC2		11	
HRC3		11	
HRC11		25	
HRC12		25	
HRC13		25	
ANC3		-	
Echo	**Alpha (%)**	**Delay**	**Feedback**
HRC5	0.3	100	0
HRC7	0.3	100	0
HRC8	0.3	100	0
HRC15	0.3	180	0.8
HRC16	0.3	180	0.8
ANC4	-	-	-

	Audio Component				Video Component			
Test Condition	Noise Type, SNR (dB)	Chop Period (s), Rate (chop/s), Mode	Clip Multiplier	Echo Alpha (%), Delay (ms), Feedback (%)	Video Codec	Bitrate (kbps)	PacketLoss PLR	Freezing Pauses, Length (s)
HRC1	car, 15	-	-	-	H.264	16000	-	1, 2
HRC2	-	-	11	-	H.264	16000	-	1, 2
HRC3	-	-	11	-	H.265	8000	0.01	-
HRC4	-	0.02, 2, zeros	-	-	H.265	8000	0.01	-
HRC5	-	-	-	0.3, 100, 0	H.264	16000	-	1, 2
HRC6	office, 10	-	-	-	H.264	16000	-	1, 2
HRC7	-	-	-	0.3, 100, 0	H.265	8000	0.01	-
HRC8	-	-	-	0.3, 100, 0	H.264	2000	0.05	-
HRC9	office, 10	-	-	-	H.264	2000	0.05	-
HRC10	office, 10	-	-	-	H.264	800	-	3, 7
HRC11	-	-	25	-	H.264	2000	0.05	-
HRC12	-	-	25	-	H.264	800	-	3, 7
HRC13	-	-	25	-	H.265	400	0.08	-
HRC14	-	0.02, 5, zeros	-	-	H.265	400	0.08	-
HRC15	-	-	-	0.3, 180, 0.8	H.264	800	-	3, 7
HRC16	-	-	-	0.3, 182, 0.8	H.265	400	0.08	-
ANC1	-	-	-	-	-	-	-	-
ANC2	-	-	-	-	-	-	-	-
ANC3	-	-	-	-	-	-	-	-
ANC4	-	-	-	-	-	-	-	-

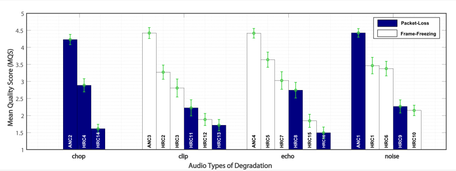
Contents

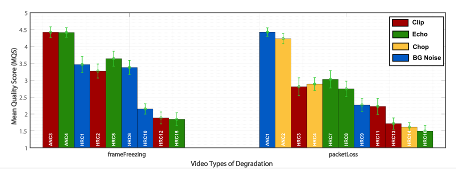


- MQS grouped by audio distortions (chop, clip, echo, and noise);
- For most HRCs, the MQS values hardly reached 3.5;
- Clip generated slightly lower quality scores, while echo HRC16 ($\alpha = 0.3$, delay = 180ms, Feedback = 0.8) received the lowest quality rating;
- Noise and Chop degradations are more sensitive to variation in parameters.

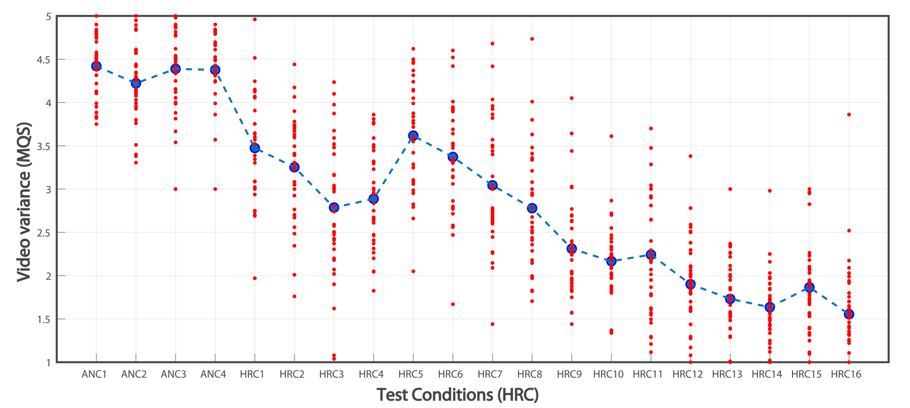


- MQS grouped by video degradations (packet-loss and frame-freezing);
- For most HRCs, the MQS hardly reaches 3.5;
- Clear difference between the MQS for packet-loss and frame-freezing distortions;
- Frame-freezing distortions seemed to have a lower impact on the perceived quality than packet-loss distortions.
- Distortion levels for Frame-freezing seemed to have a heavier impact;





- It seems that audio degradations combined with packet-loss had a stronger impact on the overall audio-visual quality.;
- For the case of audio degradation types, no particular degradation was identified as being determinant in the perceived quality.
- Regarding the video degradation types, it is clear that packet-loss has a stronger influence in the perceived quality.

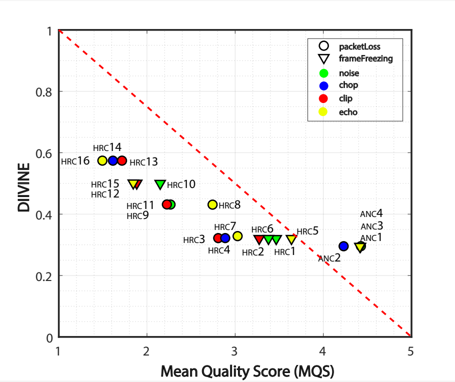


- MQS values and its respective spread of scores.
- More ‘degraded’ test conditions result in more consistent scores;

Objective Quality Comparison

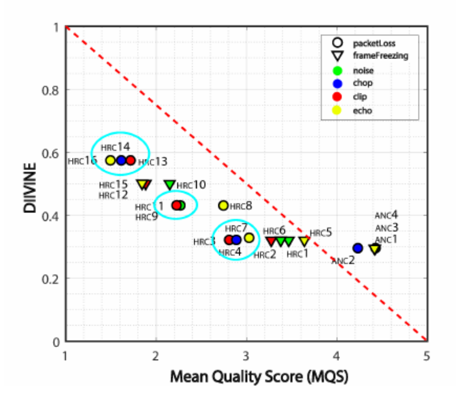
- Subjective scores correspond to the overall audio-visual quality, while the objective scores represent the predicted quality of a particular component (audio or video);
- Subjective scores are distributed on a 5-point scale (ACR), while the scores by the objective metrics are in diferent ranges, normalized to a [0,1] interval;
- The comparison between subjective and objective scores can provide interesting insights.

Objective Quality Comparison - DIIVINE



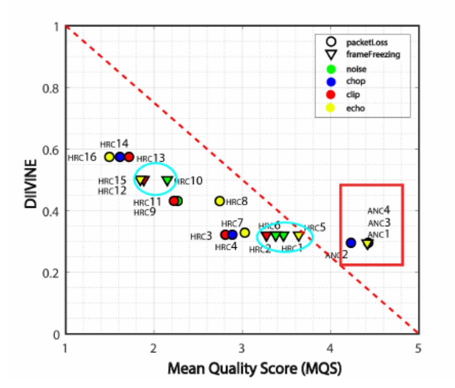
- Subjective scores versus the DIIVINE scores, organized according to the types of degradation;
- Moderate correlation;
- DIIVINE metric tend to overestimates the video quality;
- MQS values occupy most of the rating scale, DIIVINE scores are more concentrated;

Objective Quality Comparison - DIIVINE



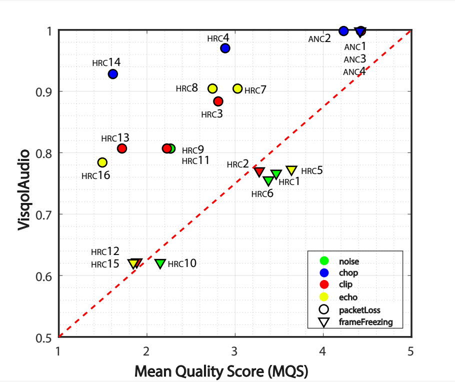
- Sequences affected by a packet-loss (HRCs 13, 14, and 16: 400 kbps, PLR = 0.08) resulted in a lower quality, according to DIIVINE;
- While sequences by frame-freezing (HRCs 1, 2, 5, and 6: 16,000 kbps P=1, L=2) were less affected;

Objective Quality Comparison - DIIVINE



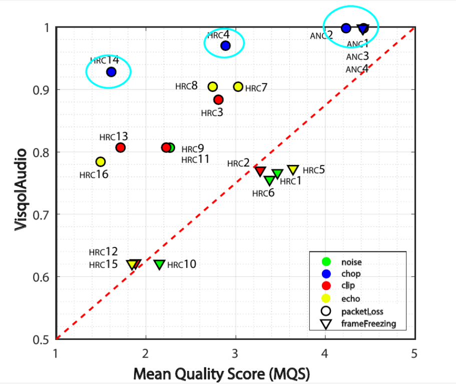
- Sequences affected by a packet-loss (HRCs 13, 14, and 16: 400 kbps, PLR = 0.08) resulted in a lower quality, according to DIIVINE;
- While sequences by frame-freezing (HRCs 1, 2, 5, and 6: 16,000 kbps P=1, L=2) were less affected;

Objective Quality Comparison - VISQOLAudio



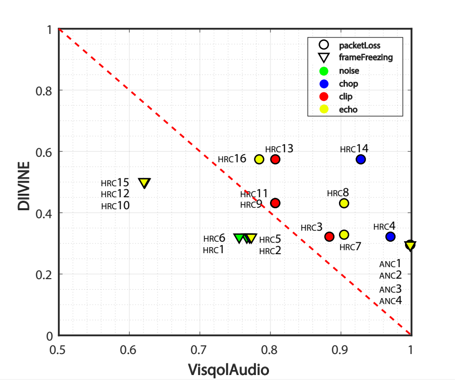
- VISQOLAudio was chosen as the audio quality metric;
- Scatter-plots of subjective audio-visual (MQS) versus VISQOLAudio scores;
- No particular pattern is observed;
- VISQOLAudio seemed to over-estimate the audio-visual quality.

Objective Quality Comparison - VISQOLAudio



- Clear difference between sequences affected by frame-freezing and packet-loss distortions;
- Similar video conditions tended to group around each other but in a lighter way compared to the previous graphs;
- Regarding the audio degradations, Chop resulted in higher quality scores.

Objective Quality Comparison

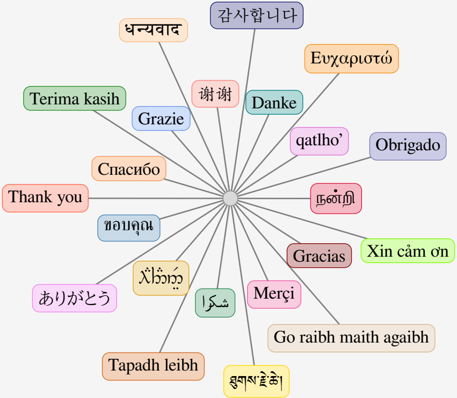


- VISQOLAudio and DIIVINE predictions were compared;
- Graph shows a disperse negative relationship between both sets of scores.

Contents

Conclusions

- Performed a subjective experiment, using the immersive methodology, with audio-visual sequences impaired with different audio and video degradations;
- Produced a database of audio-visual stimuli;
- Participants were able to distinguish the different levels of quality:
  - noise and chop degradations had a strong impact on quality;
  - packet-loss test conditions were rated lower than frame-freezing ones;
- subjective results were compared to the objective predictions of VISQOLAudio and DIIVINE scores.



Questions?

mylene@ieee.org,

http://www.ene.unb.br/mylene

http://www.ene.unb.br/mylene/databases.html