https://doi.org/10.2352/ISSN.2470-1173.2019.10.IQSP-324

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

***Helard Becerra Martinez*[*], and Mylène C.Q. Farias*[*+] ;**
[*] **Department of Computer Science,** [+] **Department of Electrical Engineering; University of Brasília, Brasília, Brazil***

Abstract

This work presents the results of a psycho-physical experiment in which a group of forty (40) human participants rated the overall quality of a set of 40 high-definition audio-visual sequences. These audio-visual sequences were impaired with audio and video types of distortions commonly encountered in an Internet-based transmission scenario. More specifically, Packet-Loss and Frame Freezing distortions were added to the video component, while Background noise, Chop, Clipping, and Echo distortions were added to the audio component. Our goal was to study how audio and visual degradations interact with each other and with the content to produce the overall audio-visual quality. An immersive experimental methodology was used to obtain more accurate observer scores. Preliminary results show that the audio and video degradations interact with each other to produce the overall audio-visual quality. For different types of audio degradations, the Clip degradation obtained slightly lower quality scores. Similarly, for the different video degradations, Frame-freezing distortions were rated higher. Also, when audio degradations were combined with Packet-loss, they had a stronger impact on the audio-visual quality.

Introduction

The area of multimedia quality assessment is a multi-disciplinary area, which combines knowledge from several domains, such as psychology, physiology, image and audio signal processing. Although the specific area of Visual Quality is fairly mature [1, 2, 3], there are still several challenges to be solved in the broader area of multimedia quality. In particular, as pointed out by Pinson et al. [4], the issue of simultaneously measuring the quality of multimedia contents (e.g. video, audio, and text) is still an open problem. In the simpler case of audio-visual content, some work has been done on trying to understand audio-visual quality, what resulted in a couple subjective models [5, 6] and a few audio-visual objective quality metrics [7, 8, 9, 10]. But, so far, few works have studied the interaction between different audio and video components [11, 12, 13], a research topic that has become very relevant given the popularity of audio-visual content.

A scenario that is particularly important is the IP-based transmission scenario. Typically, in this scenario, distortions introduced in both audio and video components affect the overall audio-visual quality, but in different ways. Additionally, these cross-modal distortions may interact with the corresponding content and with each other, making the quality model more complex [10]. Therefore, given the level of difficulty of this task, perceptual studies (e.g., psychophysical experiments) that help researchers understand the issues that affect the audio-visual quality, including the interaction of audio and video components, are of great interest to the multimedia community.

The main goal of this work is to study the impact that combinations of audio and visual degradations have on the perceived quality of audio-visual signals. With this goal, we performed a psycho-physical experiment to estimate the overall quality of audio-visual sequences containing combinations of audio-only and video-only degradations. We used an immersive experimental methodology [14] to reduce user fatigue, produce a more realistic scenario and, as a consequence, obtain robust quality scores. Considering the limited number of databases that contain audio-visual content with realistic degradations and the associated quality scores, the second objective of this work is to build a large audio-visual database and make this database available for the researcher community.

This work has two main contributions. The first is the construction of an audio-visual dataset, containing sequences with combinations of audio and video distortions and their corresponding quality scores. The dataset is composed of a diverse audio and video content, which is hard to find in the literature. The second contribution of this work is the study of the interactions of the audio and video distortions to produce the overall audio-visual quality. Although there are similar works in the literature [11, 13], most of them study how the audio and video quality interact to compose the overall quality. In this work, we focus on the common audio and video distortions introduced by transmission and compression procedures. We also present two types of transmission scenarios (UDP and TCP modeled) and several distinct audio degradations commonly encountered on a VoIP scenario, which are not often discussed in the literature.

The remainder of this document is divided as follows. First, the details of the experimental setup for the immersive experiment is presented. Next, experimental results are presented and analyzed. Then, subjective responses are compared with objective scores gathered using computational metrics. Finally, conclusions and future work are discussed.

Experimental Methodology

This experiment was designed using the Immersive Methodology proposed by Pinson [14]. This methodology tries to capture a more accurate response of the human perceived quality by presenting signals in a more natural scenario, i.e., common user consumption conditions. The immersive method has been successfully applied in several studies and the consistency of the data gathered proof it as a reliable method for this type of experiments [15, 16]. The present experiment was designed following two particular aspects of this methodology: length and content of stimuli.

Traditional methods [17, 18] restrict the stimuli duration to either 8 or 10 seconds, which might result in “artificial” sequences

Figure 1: Sixteen (16) sample frames out of the forty (40) original videos used in the subjective experiment.

with no clear idea or message transmitted. According to the immersive method, longer stimuli (e.g., 30-60 seconds) will engage participants and encourage a media-consumption real scenario. Similarly, low diversity of stimuli content is commonly used by traditional methods. Source stimuli (SRC) are later processed at a number of Hypothetical Reference Circuit (HRC) leading to large stimuli sets of repeated content. Such design tend to result on tiring experimental sessions, leading to content memorization by participants and therefore, affecting the human perceived quality responses. The immersive method tries to avoid these kinds of issues by presenting each content stimuli only once, i.e., participants will rate all HRC from the processed stimuli pool but they will not rate repeated content (SRC) from it.

Source stimuli and test conditions

A set of forty (40) high-definition audio-visual sequences were considered as source stimuli for this experiment. In order to maintain a common spatial and temporal resolution for the audio and video components, we pre-processed the videos. The video component was set to a spatial resolution of 1280 x 720 (720p), a temporal resolution of 30 frames per second (fps), and a 4:2:0 color format. Regarding the audio component, the sampling frequency was set to 48 kHz and the bit-depth was set to 16 bits.

Regarding the content, only sequences considered as "relevant" were included on the experiment. In this context, a "relevant" sequence will refer to a type of content typically consumed by a common user. The main idea was to have several categories of media entertainment. Among the desired categories, we can list sports, movies, interviews, graphic animations, and live (and studio) music. Figure 1 shows sixteen (16) sample frames of the entire dataset of 40 source videos. Abrupt cuts were avoided in order to guarantee that each sequence transmits a complete idea; this resulted on a varying length set of sequences. The sequence length varied between 19 and 62 seconds, with an average duration of 36 seconds.

A large stimuli pool was built by processing the original dataset. To generate the test stimuli pool, we introduced audio and video distortions in the audio and video components, respectively, of the original sequences. The video distortions were Bitrate compression, Packet-Loss, and Frame-Freezing. The video stimuli was compressed using H.264 and H.265 video codecs, with varying the bitrates. With respect to Packet Loss and Frame-Freezing distortions, since these types of distortions do not occur simultaneously, the videos either contained one or another type of distortion. The Packet-loss distortions were generated by dropping packets from the bitstream at different rates (PLR), while the Frame freezing distortions were generated by inserting pauses with different lengths. The test conditions were organized to produce a set of 16 Hypothetical Reference Circuits (HRCs). Table 1 shows the parameters and types of degradations of each HRC.

With respect to the audio component of the test stimuli, four (4) common streaming audio degradation types were introduced: Background noise, Chop, Clip, and Echo. These types of degradations, along with the insertion procedure, were inspired by the TCD-VoIP dataset [19]. The TCD-VoIP dataset includes some common degradations encountered in a voice over IP transmission. Degradations are considered as "platform-independent" as they are not influenced by the codec, hardware, or network in use. For this experiment, a sample of the test conditions used by the TCD-VoIP dataset was selected and inserted to the audio component of the original sequences. For each type of distortion (noise, chop, clip, and echo), two test conditions were selected and distributed along the 16 HRC arrangements. Additionally, 4 test conditions (ANC) were included as anchors. Table 1 shows the details of the HRCs and their corresponding parameters.

Altogether, 40 source stimuli were processed at 20 different test conditions (including 4 anchor conditions). This resulted in 800 different audio-visual sequences with different audio and video distortions. It is important to mention that, for each test session, the participant was presented with only 40 test stimuli of the 800 test sequences, as recommended by the immersive method.

Equipment and procedure

The experiment was conducted at the University of Brasília (UnB), in a recording studio of the Núcleo Multimedia e Internet (NMI) of the Department of Engineering (ENE). Sound isolation was guaranteed during the experiment and only one participant was allowed during each experimental session. Hardware equipment consisted of a desktop computer, an LCD monitor, a set of earphones, and a dedicated sound card to provide subjects with an ideal sound experience. Detailed specifications of the equipment are presented in Table 2.

Subjects were seated maintaining a distance of three screen heights (3H) between their eyes and the monitor screen, as it is recommended by the International Telecommunications Union on BT.500.1 [17]. This experiment was conducted with 42 participants (16 female and 26 male). Subjects were volunteers from the University of Brasília, most of them were graduate students from the Computer Science and Electrical Engineering Departments. Not having a critical hearing or vision impairment was a pre-requirement for participants, additionally, the use of glasses or contact lenses was requested if needed to watch TV.

A single experimental session was divided into three subsessions: display, training, and main sessions. The display session consisted of presenting a set of short clips containing all test conditions (HRCs) to the participant. Test conditions included all degradation levels for both audio and video distortions. The main purpose of this session was to give the participant an idea of the quality range considered in the experiment. Once the session was completed, participants were asked if they have noticed quality

Table 1: Coding parameters and types of degradations of the audio and video component of each HRC of the dataset.

	Audio Component				Video Component			
	Noise	Chop	Clip	Echo	Video Codec	Bitrate (kbps)	PacketLoss	Freezing
HRC	Type, SNR (dB)	Period (s), Rate (chop/s), Mode	Multiplier	Alpha (%), Delay (ms), Feedback (%)			PLR	Pauses, Length (s)
HRC1	car, 15	-	-	-	H.264	16,000	-	1, 2
HRC2	-	-	11	-	H.264	16,000	-	1, 2
HRC3	-	-	11	-	H.265	8,000	0.01	-
HRC4	-	0.02, 2, zeros	-	-	H.265	80,00	0.01	-
HRC5	-	-	-	0.3, 100, 0	H.264	16,000	-	1, 2
HRC6	office, 10	-	-	-	H.264	16,000	-	1, 2
HRC7	-	-	-	0.3, 100, 0	H.265	8,000	0.01	-
HRC8	-	-	-	0.3, 100, 0	H.264	2,000	0.05	-
HRC9	office, 10	-	-	-	H.264	2,000	0.05	-
HRC10	office, 10	-	-	-	H.264	800	-	3, 7
HRC11	-	-	25	-	H.264	2,000	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	-	-	-	-	H.264	64,000	-	-
ANC2	-	-	-	-	H.265	32,000	-	-
ANC3	-	-	-	-	H.264	64,000	-	-
ANC4	-	-	-	-	H.265	32,000	-	-

Table 2: Equipment specifications

Equipment	Technical Details
Monitor	Samsung SyncMaster P2370
	Resolution: 1,920x1,080; Pixel-response rate: 2ms;
	Contrast ratio: 1,000:1; Brightness: 250cd/m2
Earphones	Sennheiser Hd 518 Headfone
	Impedance: 50 Ohm; Sound Mode: Stereo;
	Frequency response: 1426,000Hz;
Sound Card	Asus Xonar DGX 5.1

differences between the displayed clips.

The next session was the training session. The goal of this session was that the participant got used to the experimental protocol and with the rating procedure used in the main session. In this session, we presented sample sequences to the participants, which contained different levels of distortion (both audio and video). After each sequence was displayed, a rating scale appeared on the screen and the participant was requested to rate the sequence using a five-point Absolute Category Rating (ACR) scale, ranging from 1 to 5. The rating scale was labeled (from 1 to 5) as “Bad”, “Poor”, “Fair”, “Good”, and “Excellent”.

Finally, during the main session, the actual experiment was performed, following the same procedure used in the training session. The experimental methodology was single stimulus, with each test sequence being played only once. As mentioned earlier, four anchor sequences were included in the dataset. Sequences were played in a random and rated using the ACR scale. In total, considering all three sessions, a single experiment lasted around 50 minutes on average. To avoid fatigue, participants could take a small break in the middle of the experiment.

Experiment Results

Responses collected from participants in this type of experiment are known as subjective scores. For traditional methods, the mean opinion score (MOS) associated with a single test sequence is obtained by averaging all scores given by all participants for that particular sequence. In our experiment, the Mean Quality Score (MQS) per-HRC is obtained by averaging the quality scores, given by all participants, for a particular j-th HRC:

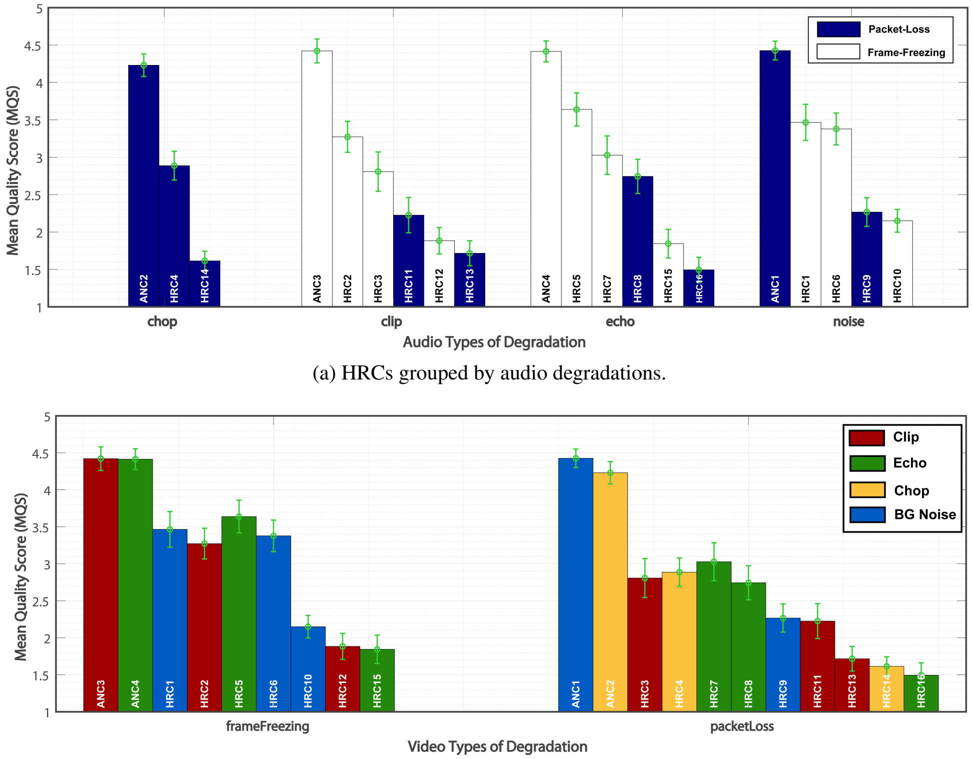
$$\mathrm{MQS(j)} = \frac{1}{n} \cdot \sum_{i=0}^{n} QS(i,j), \tag{1}$$

where n is the total number of subjects (in our case $n = 12$) and QS(i,j) is the quality score given by the i-th subject to the j-th HRC test sequence. In other words, MQS(j) gives the average quality score for the j-th HRC, measured over all subjects and originals.

Figure 2 presents the MQS values collected from the subjective experiment. In Figure 2 (a) the MQS values are grouped according to the audio distortions (chop, clip, echo, and noise), meanwhile in Figure 2 (b) the values are grouped according to the video degradations (packet-loss and frame-freezing). It can be observed in this figure that most HRCs obtained quality scores equal or below 3.5, while the anchors sequences (ANC) obtained quality scores well above 4. Considering the different types of audio degradations, Clip degradations obtained slightly lower quality scores on average, while Echo test condition HRC16 received the lowest quality score. Additionally, by observing the gaps between the distortion levels for Clip and Echo distortions, we noticed that the differences between neighboring HRCs were roughly constant, while the differences between neighboring HRCs for Noise and Chop seemed more irregular. This might suggest that Noise and Chop degradations were more sensitive to variations, i.e., varying the distortion level for these distortions had a higher impact on the perceived quality.

In Figure 2 (b), where MQS scores were organized according to the different types of video degradations, we notice that there is a clear difference between the MQS values obtained for the Packet-loss and Frame-freezing distortions. On average, Frame-freezing distortions seemed to have a lower impact on the perceived quality than Packet-loss distortions. However, by observing the gaps between both types of distortions, variations of Frame-freezing distortion levels seemed to have a heavier impact on the perceived quality. In other words, varying the levels of distortion for Frame-freezing produced a more pronounced drop of quality, when compared to a variation in Packet-loss distortion.

For the case of audio degradations, no particular degradation was identified as having a determinant effect on the perceived quality. As already mentioned, for the case of video degradations, Packet-loss had a stronger influence on the perceived audio-visual quality. Therefore, in terms of combined degradations, audio degradations combined with Packet-loss had a stronger impact on the overall audio-visual quality.



(a) HRCs grouped by audio degradations.

(b) HRCs grouped by video degradations.

Figure 2: Mean Quality Score (MQS) for the different combinations of audio and video degradations (Table 1 describes each HRC).

Figure 3 presents the MQS values obtained for each of the HRCSs, along with the single user scores. It can be observed that for more ‘degraded’ HRC (see Table 1), the results are more consistent, i.e., the spread of points is smaller. But, for HRCs that received a MQS value around the center of the scale, the scores provided by participants varied more, resulting in a larger standard deviations around the average value.

Objective Quality Comparison

We compare the subjective scores with the objective results gathered from one audio and one video quality metrics. Naturally, the subjective scores correspond to the overall audio-visual quality, while the quality scores predicted by the objective metrics represent the quality of a particular component (audio or video). Also, it is worth pointing out that the subjective scores are distributed on a five-point rating scale (ACR), while the scores predicted by the objective metrics do not have the same range, which might lead to scale calibration bias. Despite these issues, the comparison between subjective and objective scores can provide interesting insights concerning the predicted quality and their interaction with the overall audio-visual perceived quality.

The DIIVINE quality metric [20] was selected to predict the quality of the video component of the stimuli. Figure 4 depicts the scatter-plots of the subjective scores versus the corresponding DIIVINE scores, organized according to the types of degradation. In general terms, and independent if it is an audio or a video degradation, the scatter-plots presented a moderate negative correlation between the subjective audio-visual (MQS) and the DIIVINE scores. It seems that the DIIVINE metric tended to overestimate the video quality of sequences, since most points fall below the red line in the graph (this being interpreted as better quality). While MQS values occupied most of the rating scale (1 to 5), DIIVINE scores were concentrated on the middle of their scale (0 to 1). Despite this characteristic, DIIVINE scores varied along the MQS values, showing a good consistency.

Figure 4 shows that sequences affected by Packet-loss degradations (HRCs 13, 14, and 16) resulted in a lower quality, according to the DIIVINE metric. The same graph suggests that sequences with a Frame-freezing type of degradation (HRCs 1, 2, 5, and 6) were less affected in terms of quality. Naturally, regarding the audio distortions, no particular behavior was observed in terms of a higher or lower quality for a specific audio degradation. However, it can be observed that video degradations tend to group around similar conditions. This tendency is only broken for two cases that correspond to Noise and Chop audio degradations (HRCs 10 and 8), which suggests an influence of audio distortions on the perceived audio-visual quality.

VISQOLAudio was chosen as the audio quality metric [21]. Figure 5 depicts the scatter-plots of the subjective audio-visual quality scores (MQS) versus the VISQOLAudio scores, organized according to the audio and video types of degradation. In general terms, and considering that this comparison is made between audio and audio-visual scores, no particular pattern was observed. VISQOLAudio also seemed to overestimate the quality for most conditions (most marks fall above the red line), which is expected since only the audio component is being measured.

In Figure 5 it can be observed a clear difference between sequences affected by Frame-freezing and Packet-loss distortions.

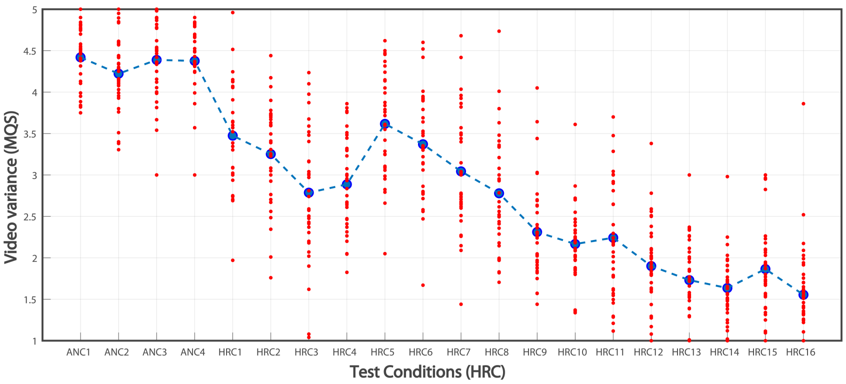


Figure 3: Mean Quality Score (MQS), and its respective spread of scores, for the different Hypothetical Reference Circuit (HRC) degradations (see Table 1).

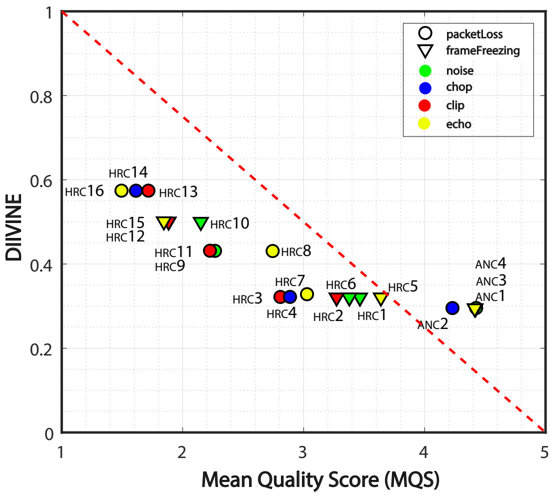


Figure 4: Scatter plot of audio-visual subjective scores (MQS) versus video objective scores (produced by DIIVINE).

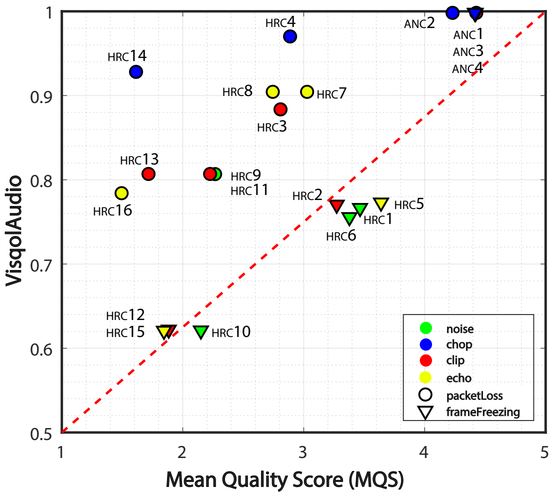


Figure 5: Scatter plot of audio-visual subjective scores (MQS) versus audio objective scores (produced by VISQOLAudio).

Again, similarly, video conditions tended to group around each other, but not as 'strongly' as it was seen in Figure 4. Regarding the type of audio degradations, Figure 5 shows that Chop sequences got higher quality scores.

Finally, both VISQOLAudio and DIIVINE scores were compared. Figure 6 depicts a scatter-plot of these scores, organized by the types of audio and video degradations. The graph shows a disperse negative relationship between both sets of scores. It can be observed that scores remained spread in the middle of the rating scale. It can be noticed that frame-freezing conditions (HRCs 10, 12, and 15) presented lower audio and video quality predictions.

Conclusions

This work presented a subjective quality experiment conducted using the immersive methodology. The experiment presented audio-visual sequences impaired with different audio and video types of degradations. For the video component, sequences were impaired with two types of degradations: Packet-loss and Freezing-frames. As for the audio component, four common streaming audio degradation types were considered: Noise, Chop, Clip, and Echo. A large dataset, consisting of 800 audio-visual sequences with both audio and video distortions was made public. Given the difficulty of finding public datasets with this type of characteristics, it is expected that the present work will be a good contribution to the area and encourage further studies in audio-visual quality assessment.

Experimental results suggests that participants were able to distinguish between the different levels of quality for each type of degradation. For the particular case of Noise and Chop degradations, it could be observed that variations on the levels of these distortions had a strong impact on the perceived quality. As for the video degradations, Frame-freezing test conditions were rated with higher quality when compared to Packet-loss test conditions. Similarly to Noise and Chop degradations, varying the level of distortion of Frame-freezing sequences presented a strong impact on the perceived quality.

Finally, subjective results were compared to the objective predictions of an audio (VISQOLAudio) and a video (DIIVINE) quality metric. By comparing the audio-visual MQS and the DIIVINE predictions, we noticed a tendency to overestimate video

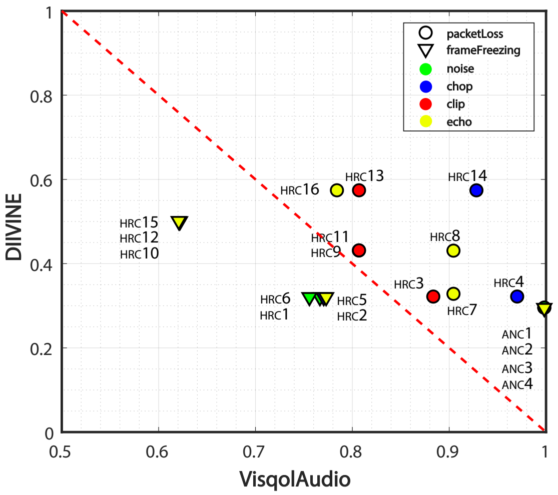


Figure 6: Scatter plot of audio objective scores (prduced by VISQOLAudio) versus video objective scores (produced by DIIVINE).

quality. It was also observed that DIIVINE scores varied along MQS range, suggesting a consistency on results predicted by this video metric. A more detailed analysis showed that Packet-loss conditions (HRCs 13, 14, and 16) were rated with lower quality, compared to Frame-freezing test conditions. It was also noticed that video test conditions tended to group around each other, this tendency was only broken for conditions where the audio component was affected by Noise and Chop degradations. This might suggest that, although the video component is the strongest influential factor, certain audio types of degradation play an important role in terms of the overall audio-visual quality.

A similar comparison of the audio-visual MQS and the audio predictions from VISQOLAudio showed, again, that the objective metric overestimates the perceived audio-visual quality. It was also observed that Chop sequences received a higher quality prediction. As observed with DIIVINE results, video test conditions also tended to group around each other, but in a lighter way. A final comparison between DIIVINE and VISQOLAudio results showed that most predictions fall in the middle of the rating scale, it was also observed that some Frame-freezing conditions (HRCs 10, 12, and 15) received low audio and video quality predictions.

Acknowledgments

This work was supported in part by the Conselho Nacional de Desenvolvimento Cientfico e Tecnológico (CNPq), the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES), and the University of Brasília (UnB).

References

[1] Chikkerur, S., Sundaram, V., Reisslein, M., and Karam, L. J., "Objective video quality assessment methods: A classification, review, and performance comparison," *IEEE transactions on broadcasting* **57**(2), 165 (2011).

[2] *Perceptual visual quality metrics: A survey*, **22**(4) (2011).

[3] *Visual quality assessment algorithms: what does the future hold?*, **51** (February 2011).

[4] Pinson, M., Ingram, W., and Webster, A., "Audiovisual quality components," *IEEE Signal Processing Magazine* **6**(28), 60–67 (2011).

[5] Soh, K. and Iah, S., "Subjectively assessing method for audiovisual quality using equivalent signal-to-noise ratio conversion," *Trans. Inst. Electron., Inform. Commun. Eng. A* **11**, 1305–1313 (2001).

[6] Hands, D. S., "A Basic Multimedia Quality Model," *Multimedia, IEEE Transactions on* **6**(6), 806–816 (2004).

[7] Garcia, M. and Raake, A., "Impairment-factor-based audio-visual quality model for iptv," in *Int. Workshop on Quality of Multimedia Experience (QoMEx), 2009.*, 1–6, IEEE (2009).

[8] Yamagishi, K. and Gao, S., "Light-weight audiovisual quality assessment of mobile video: Itu-t rec. p.1201.1," in *IEEE Int. Workshop on Multimedia Signal Processing (MMSP), 2013*, 464–469.

[9] Martinez, H. A. B. and Farias, M. C., "Combining audio and video metrics to assess audio-visual quality," *Multimedia Tools and Applications* , 1–20 (2018).

[10] Fernández, I. B. and Leszczuk, M., "Monitoring of audio visual quality by key indicators," *Multimedia Tools and Applications* **77**(2), 2823–2848 (2018).

[11] You, J., Reiter, U., Hannuksela, M., Gabbouj, M., and Perkis, A., "Perceptual-based quality assessment for audio–visual services: A survey," *Signal Processing: Image Communication* **25**(7), 482–501 (2010).

[12] Belmudez, B. and Möller, S., "Audiovisual quality integration for interactive communications," *EURASIP Journal on Audio, Speech, and Music Processing* **2013**(1), 24 (2013).

[13] Akhtar, Z. and Falk, T. H., "Audio-visual multimedia quality assessment: A comprehensive survey," *IEEE Access* **5**, 21090–21117 (2017).

[14] Pinson, M., Sullivan, M., and Catellier, A., "A new method for immersive audiovisual subjective testing," in [*Proceedings of the 8th International Workshop on Video Processing and Quality Metrics for Consumer Electronics (VPQM)*], (2014).

[15] Garcia, M.-N., Dytko, D., and Raake, A., "Quality impact due to initial loading, stalling, and video bitrate in progressive download video services," in *IEEE Int. Workshop on Quality of Multimedia Experience (QoMEX), 2014*, 129–134.

[16] Robitza, W., Garcia, M. N., and Raake, A., "At home in the lab: Assessing audiovisual quality of http-based adaptive streaming with an immersive test paradigm," in [*Int. Workshop on Quality of Multimedia Experience (QoMEX), 2015*.

[17] [*ITU-T Recommendation BT.500-8: Methodology for the subjective assessment of the quality of television pictures*] (1998).

[18] ITU-R, "Recommendation P.911 : Subjective audiovisual quality assessment methods for multimedia applications," (1998).

[19] Harte, N., Gillen, E., and Hines, A., "TCD-VOIP, a research database of degraded speech for assessing quality in VOIP applications," in *IEEE Int. Works. on Quality of Multimedia Experience (QoMEX), 2015*.

[20] Zhang, Y., Moorthy, A. K., Chandler, D. M., and Bovik, A. C., "C-diivine: No-reference image quality assessment based on local magnitude and phase statistics of natural scenes," *Signal Processing: Image Communication* **29**(7), 725–747 (2014).

[21] Hines, A., Gillen, E., Kelly, D., Skoglund, J., Kokaram, A., and Harte, N., "Visqolaudio: An objective audio quality metric for low bitrate codecs," *The Journal of the Acoustical Society of America* **137**(6), EL449–EL455 (2015).

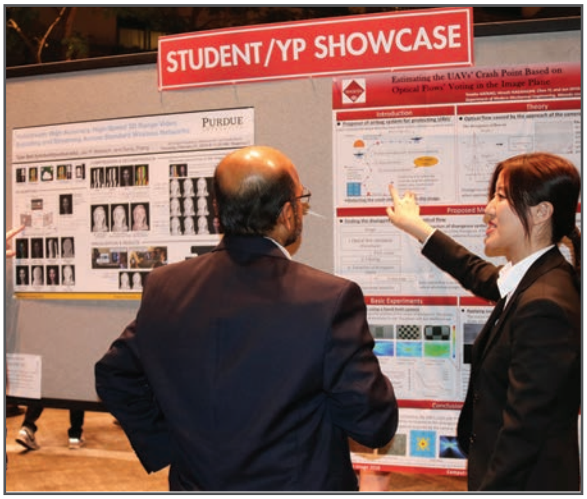
JOIN US AT THE NEXT EI!

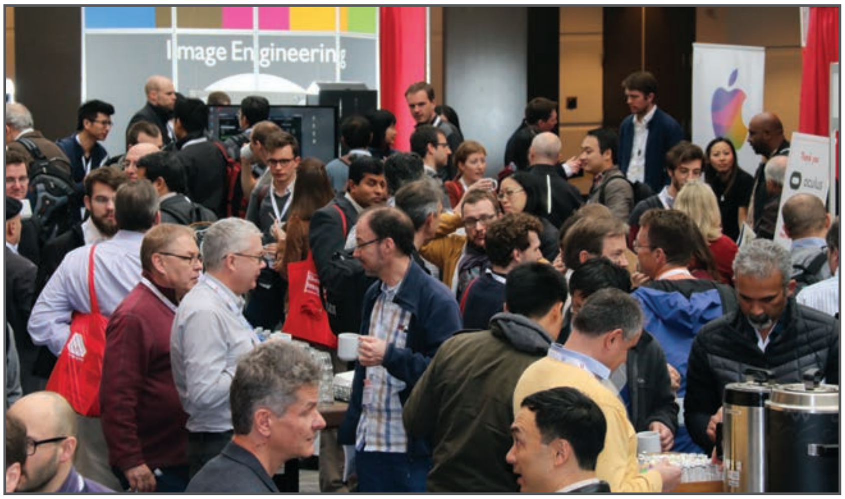
IS&T International Symposium on

Electronic Imaging

SCIENCE AND TECHNOLOGY

Imaging across applications . . . Where industry and academia meet!





AR is a fundamental and transformative interface and platform.

20+ years from my initial working demo to today's AR displays

Optimistic that it won't take another 20 to reach ubiquitous consumer AR!

My first working AR prototype

Today's platforms

Optimis this is << 20

Ubiquitous consumer AR

• SHORT COURSES • EXHIBITS • DEMONSTRATION SESSION • PLENARY TALKS •
• INTERACTIVE PAPER SESSION • SPECIAL EVENTS • TECHNICAL SESSIONS •

www.electronicimaging.org

IS&T
imaging.org