Audiovisual Quality Evaluation of Low-Bitrate Video

Stefan Winkler[a] and Christof Faller[b]

[a] Genista Corporation, Rue du Théâtre 5, 1820 Montreux, Switzerland
[b] Audiovisual Communications Lab, Ecole Polytechnique Fédérale de Lausanne (EPFL), Switzerland

ABSTRACT

Audiovisual quality assessment is a relatively unexplored topic. We designed subjective experiments for audio, video, and audiovisual quality using content and encoding parameters representative of video for mobile applications. Our focus were the MPEG-4 AVC (a.k.a. H.264) and AAC coding standards.

Our goals in this study are two-fold: we want to understand the interactions between audio and video in terms of perceived audiovisual quality, and we use the subjective data to evaluate the prediction performance of our non-reference video and audio quality metrics.

Keywords: Audiovisual quality metrics, MPEG-4 AAC, AVC/H.264, 3GPP, ACR

1. INTRODUCTION

Video quality assessment[1] has become rather well established by now, as evidenced by the number of research publications available, as well as the collaborative efforts of the Video Quality Experts Group (VQEG).[2, 3] Standards for subjective assessment[4, 5] have been around for many years. A variety of products are already available on the market, and the ITU recently recommended several full-reference quality metrics for TV applications.[6, 7]

Audio and especially speech quality assessment have an even longer history. There are several standards for subjective tests.[8, 9] Additionally, speech and audio quality metrics have been standardized in the form of PESQ[10] and PEAQ,[11] respectively.

Audiovisual quality, however, is an entirely different matter. There has been some work in the past,[12–15] but it mainly concerns the subjective side, and nothing has been published so far with respect to low bitrates and today's codecs. This area will be the focus of the present paper.

Mobile applications are our principal interest here. They are characterized by a specific set of requirements that include low bitrates, small frame sizes, and low frame rates. Furthermore, the content is viewed at short distance on a small LCD screen with a progressive display. For our experiments, we selected source material covering a representative set of content. The source clips were encoded with codecs well-suited for mobile applications, namely MPEG-4 AVC,[16] also known as H.264,[17] traditional MPEG-4[18] and H.263[19] for video as well as MPEG-4 AAC[20] for audio, using appropriate coding parameters. The optimum choice of coding parameters and audio/video bitrate ratios for achieving the best quality with these codecs is discussed in detail elsewhere.[21]

Subjective ratings were obtained for the resulting test clips for audio-only, video-only and audiovisual presentations using the Absolute Category Rating (ACR) methodology defined by ITU-T Recommendations P.910 and P.911.[5, 22] We analyze the results of these experiments with respect to inter-subject variability. Furthermore, we study the influence of perceived audio and video quality on audiovisual quality. Finally, we combine non-reference metrics for video and audio to compute predictions of perceived quality. We show that their predictions match well with the subjective ratings obtained.

The paper is organized as follows. Section 2 introduces the source material, the simulation environment, and the test conditions used to produce the test clips. In Section 3 we describe the subjective assessment method, the presentation of the sequences, and the viewing conditions. In Section 4 we analyze the data obtained in the subjective experiments and discuss the interactions between audio and video quality. Finally, we evaluate the predictions of audio and video quality metrics in Section 5.

E-mail of corresponding author: stefan.winkler@genista.com

2. TEST MATERIAL

2.1. Source Clips

The content of the source clips was chosen to be representative of a typical scenario for viewing audiovisual content on a mobile device. The source material comprises 6 short clips of about 8 seconds each and covers a wide range of coding complexity. The visual content and the soundtrack of these scenes is summarized in Table 1. The video source material was originally in TV format; for our tests we de-interlaced and downsampled it to QCIF frame size (176 × 144 pixels). The audio source material was 16-bit PCM stereo sampled at 48 kHz.

Table 1: Content of test scenes.

Scene	Name	Video	Audio	Duration
A	Buildings	slow horizontal pan across a city skyline, followed by a vertical pan up a building facade	orchestral background music	7.48 sec.
B	Conversation	camera switching between head-and-shoulders shots of a woman and a man talking	male and female voices	8.36 sec.
C	Football	American football scene from VQEG;[2] high motion	crowd cheering and chanting; female commentator	7.60 sec.
D	Music video	music video clip; high motion	rock music with vocals	8.08 sec.
E	Trailer 1	action movie trailer; scene cuts and high motion	theme music and voice-over	8.84 sec.
F	Trailer 2	romance movie trailer with credits; scene cuts	theme music and voice-over	8.08 sec.

2.2. Test Conditions

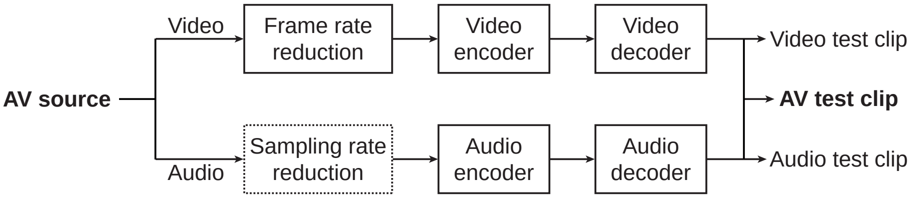

Figure 1: Test material creation.

The setup for the creation of the test material is illustrated in Figure 1. Before encoding, the frame rate of the source video clips was reduced to 8 fps or 15 fps using VirtualDub.*

Our codec selection was principally determined by the 3GPP† file format.[23] It is of particular interest for packet-switched video streaming in 3G networks. It can contain bitstreams of H.263,[19] traditional MPEG-4,[18] and since its latest release even MPEG-4 AVC/H.264[16, 17] for video as well as MPEG-4 AAC[20] and different types of AMR (Advanced Multi-Rate) for audio.

For the video track, we chose all three codecs supported by the 3GPP file format. QuickTime Pro version 6.5 was used for H.263 and MPEG-4 encoding; it does not currently support H.264. Therefore, the JM reference software‡ version 8.5 was used for H.264 encoding (baseline profile), even though its output may not be strictly 3GPP-compliant.

Unfortunately, the QuickTime codecs for MPEG-4 and especially H.263 did not produce substantial quality variations within the bitrate range of interest.[21] Viewers would have been unable to discern the quality of the different test clips. Furthermore, the QuickTime encoders did not achieve the target bitrates at the low end of the range. The H.264 JM reference encoder does not have these problems. We therefore decided to use H.264 for most test conditions.

---

* VirtualDub is available at http://www.virtualdub.org/

† 3rd Generation Partnership Project, see http://www.3gpp.org/

‡ The JM reference software is available at http://bs.hhi.de/~suehring/tml/

Table 2: Video test conditions.

Condition	Codec	Frame rate	Bitrate
1	H.264	8 fps	24 kb/s
2	H.264	8 fps	32 kb/s
3	H.264	8 fps	40 kb/s
4	H.264	8 fps	48 kb/s
5	H.263	8 fps	48 kb/s
6	MPEG-4	8 fps	48 kb/s
7	H.264	15 fps	24 kb/s
8	H.264	15 fps	32 kb/s
9	H.264	15 fps	40 kb/s
10	H.264	15 fps	48 kb/s

Table 3: Audio test conditions.

Condition	Channels	Sampling rate	Bitrate
1	mono	8 kHz	8 kb/s
2	mono	16 kHz	16 kb/s
3	mono	22 kHz	24 kb/s
4	mono	32 kHz	32 kb/s
5	mono	22 kHz	32 kb/s
6	stereo	22 kHz	32 kb/s
7	stereo	16 kHz	32 kb/s

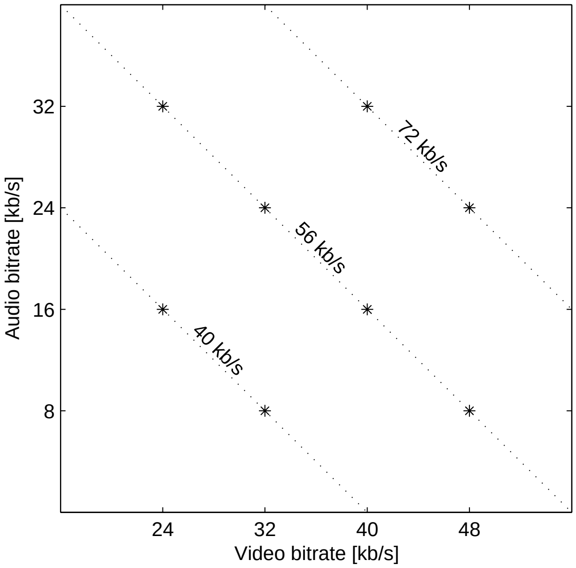

Figure 2: Audiovisual test conditions. The stars denote the video and audio bitrate combinations used in the test. The diagonal dotted lines connect points with the same total data rate (indicated by their labels).

For the audio track, we chose the MPEG-4 AAC-LC (low complexity) coding standard.[20] QuickTime Pro was again used for encoding, with the "recommended" sampling rate for each target bitrate (i.e. the sampling rate reduction depicted in Figure 1 was performed internally by the encoder).

Video conditions 1–4 from Table 2 were combined with audio conditions 1–4 from Table 3 for a total of 8 audiovisual test conditions as illustrated in Figure 2. Of particular interest is a total data rate of 56 kb/s, which can be transmitted over a typical 64 kb/s wireless link. Data rates of 40 kb/s and 72 kb/s were included as well.

3. SUBJECTIVE ASSESSMENT

3.1. Assessment Method

The experimental set-up follows ITU-T Recommendations.[5, 22] We use ACR (Absolute Category Rating), a very efficient testing methodology, where the test clips are viewed one at a time and rated independently on a discrete 11-level scale from "bad" to "excellent" (see Figure 4). The ratings for each test clip are then averaged over all subjects to obtain a Mean Opinion Score (MOS).

Our initial plan was to use hidden reference removal as proposed by some studies as well as upcoming VQEG evaluations for single stimulus tests. Hidden reference implies that the subjects are not aware of the fact that the original uncompressed clips are included in the test. The "removal" of the hidden reference is done in the analysis by subtracting each subject's score for the reference from the corresponding test clips. However, we found the quality difference between reference and compressed clips to be so large that we decided against including the reference clips in the set evaluated by the subjects.

3.2. Presentation Structure

The subjective test consisted of one session of about 40 minutes, including training. Written instructions were given to the subjects. A short training session preceded the actual test; it comprised three audiovisual clips demonstrating the extremes of expected audio, video, and audiovisual quality ranges. The subjects were allowed to adjust the volume to a comfortable level during the training session. The actual test consisted of three consecutive parts, which are listed in Table 2.

Table 2: Subdivision of subjective test session.

#	Part	Result	Conditions	Clips	Comment
1	Audiovisual quality (AVQ)	$\mathrm{MOS}_{\mathrm{AV}}$	Figure 2	48	
2	Audio-only quality (AQ)	$\mathrm{MOS}_{\mathrm{A}}$	Table 3	42	blank (gray) screen
3	Video-only quality (VQ)	$\mathrm{MOS}_{\mathrm{V}}$	Table 2	60	muted audio

The subjects were asked to rate the quality of the presentation in each case. This would allow us to relate perceived audio quality and perceived video quality with the overall audiovisual quality ratings. The order of the clips was randomized individually for each subject.

3.3. Setup

The tests were conducted in a dark and sound-insulated room. The lab setup is shown in Figure 3. The monitor used in the subjective assessments was a 17" LCD screen (Dell 1703FP), which was operated at its native resolution of 1280×1024 pixels. The video clips were displayed at their original size (QCIF) in the center of the screen, surrounded by a uniform gray background. The viewing distance was not fixed. For our test material, we found subjects to be most comfortable at a viewing distance of around 30-40 cm, which corresponds to about 8-10 times the height of the video picture in our setup.

For audio playback an external D/A converter (Emagic EMI A26) was connected to the PC. High-quality headphones (Sennheiser HD 600) were directly connected to the D/A converter.

Genista's *QualiView* software was used for the playback of the test clips. It reads the decoded clips (both video and audio) stored in uncompressed AVI format and plays them out. After each clip, the voting dialog shown in Figure 4 is presented on the screen, and the rating entered by the subject is recorded.

Figure 3: Testing lab setup.

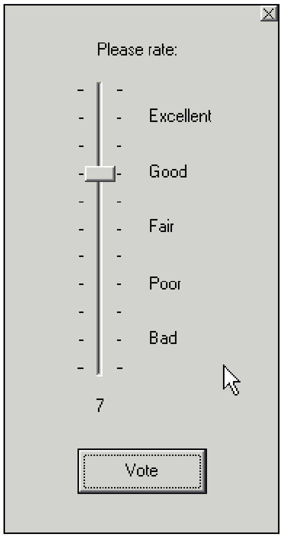

Figure 4: ACR voting dialog.

3.4. Subjects

24 subjects (6 female, 18 male) participated in the test. Their age ranged from 25 to 36 years. Some of the subjects were familiar with audio or video processing. All subjects reported that they had normal or corrected vision and normal hearing.

4. SUBJECTIVE DATA ANALYSIS

4.1. Inter-Subject Agreement

As a reliability indicator of the subjective data, the distributions of the 95%-confidence intervals and the corresponding MOS values are shown in Figure 5. The average confidence interval is a bit smaller than for Single Stimulus Continuous Quality Evaluation (SSCQE) data, but slightly larger than for Double Stimulus Impairment Scale (DSIS) data.[24] However, note that the test content and conditions were different in those tests. In general, the confidence interval sizes indicate a good agreement between subjects, despite the use of an absolute rating methodology.

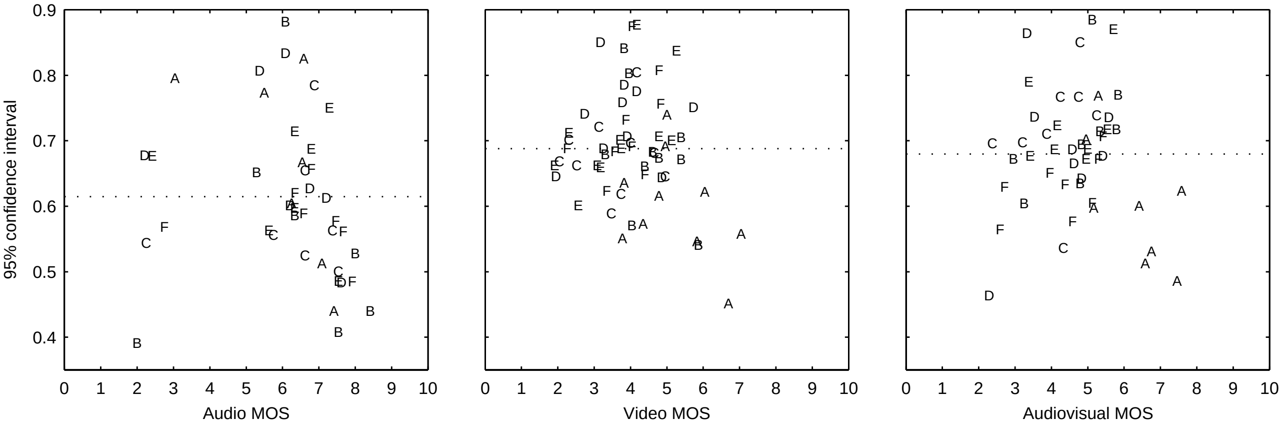

Figure 5: 95%-confidence interval size vs. MOS for the three parts of the test. The letters represent the scene descriptors from Table 1. The average confidence interval size is shown as a dotted horizontal line in each plot.

The confidence intervals are smallest for the audio-only data. However, this difference is not very significant, and we do not have sufficient data to conclude whether this is true for audio testing in general or simply due to the specific test material chosen. It is interesting to note that the confidence interval size is mainly content-dependent (and not condition-dependent), as indicated by the clustering of certain scenes in the plots.

As can be seen from the MOS distributions, subjects hesitated to use the entire range of the ACR scale, especially for the VQ and AVQ parts of the test. MOS values below 2 and above 8 are rare. For the AQ part, the perceived quality difference between condition 1 and the others leaves a noticeable gap in the middle of the range.

The plots also show the relationship between confidence interval size and MOS. In a previous study we found the agreement between observers to be highest at both ends of the MOS scale, whereas the largest confidence intervals occurred in the medium-quality regime.[25] This is also true for our data, but it is not as pronounced, primarily because of the smaller test set with fewer data points, and because subjects did not use the full range of the rating scale.

4.2. Quality Ratings

MOS_A and MOS_V values are shown as a function of bitrate and scene in Figure 6. The VQ variation with bitrate is not so large, but the scene has a big influence on perceived quality. The opposite is true for AQ; the big difference between conditions 1 and 2 (8 kb/s and 16 kb/s) is evident.

MOS_AV values are shown as a function of video bitrate in Figure 7. The optimum audio/video bit budget allocation is clearly scene-dependent; generally, 32–40 kb/s for video and 16–24 kb/s for audio produce the best AVQ at a total bitrate of 56 kb/s. As expected, a reduction of the total bitrate leads to a decrease in quality, and vice versa; the effect is most pronounced for the higher video bitrates in both cases. A more in-depth analysis of the optimal trade-off between audio and video in terms of bitrate budget and AVQ for this material is provided elsewhere.[21]

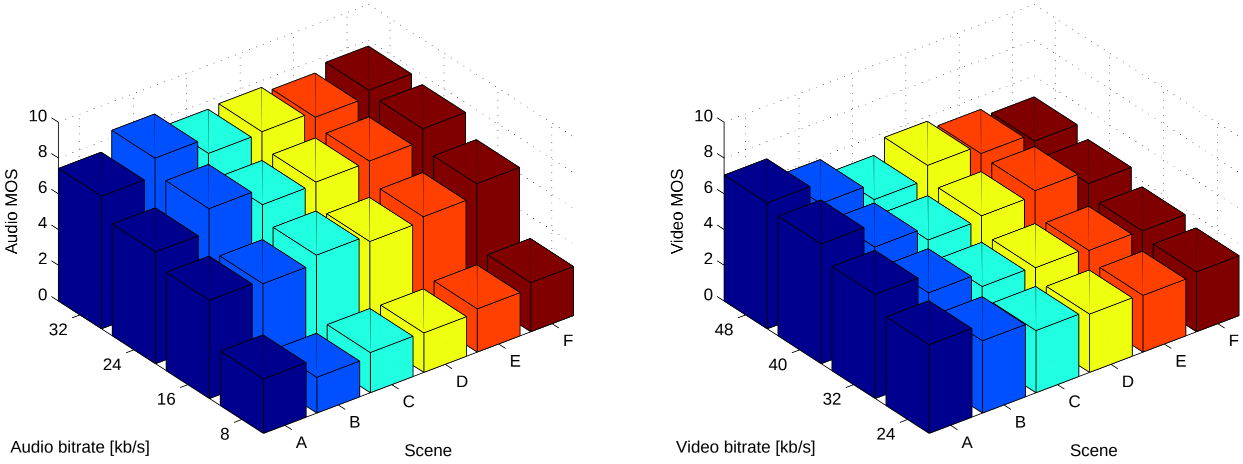

Figure 6: $\mathrm{MOS_A}$ (left) and $\mathrm{MOS_V}$ (right) as a function of bitrate and scene.

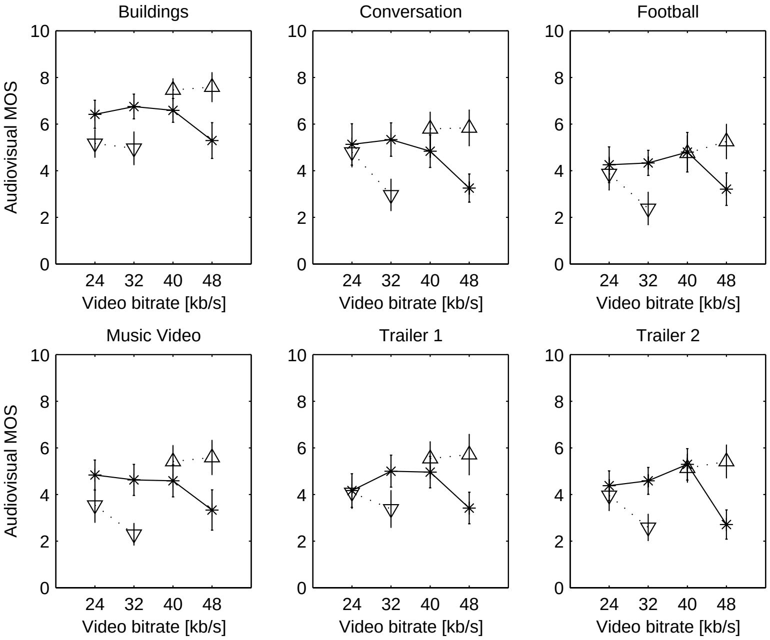

Figure 7: $\mathrm{MOS_{AV}}$ as a function of video bitrate (audio bitrate = total bitrate − video bitrate) for total bitrates of 56 kb/s (stars), 40 kb/s (downward-pointing triangles) and 72 kb/s (upward-pointing triangles).

4.3. Audio-Video Interactions

4.3.1. Principal Component Analysis

To study the influence of AQ, VQ, and the multiplicative interaction term AQ·VQ on AVQ, we carried out a principal component analysis (PCA). Four-dimensional test vectors composed of $\mathrm{MOS_A}$, $\mathrm{MOS_V}$, $\mathrm{MOS_A} \cdot \mathrm{MOS_V}$ and $\mathrm{MOS_{AV}}$ values were constructed. Each vector contains $\mathrm{MOS_A}$ and $\mathrm{MOS_V}$ obtained with the same bitrates as were used for the corresponding $\mathrm{MOS_{AV}}$ item. Prior to the PCA, the mean of the data was removed and the variance was normalized for each dimension.

Figure 8 shows the eigenvalues corresponding to the four principal components. Since 97% of the variability is explained by the first two principal components, we plot AQ, VQ, AQ·VQ as well as AVQ vectors relative to the first two principal components in Figure 9. This plot indicates that neither AQ nor VQ alone determine AVQ; both have a similarly strong influence on AVQ. However, the multiplicative term AQ·VQ is rather close to AVQ. Note that AQ and VQ are more different from AVQ than in another study.[13] This difference may be due to the small size or the low bitrates of the video clips in our test.

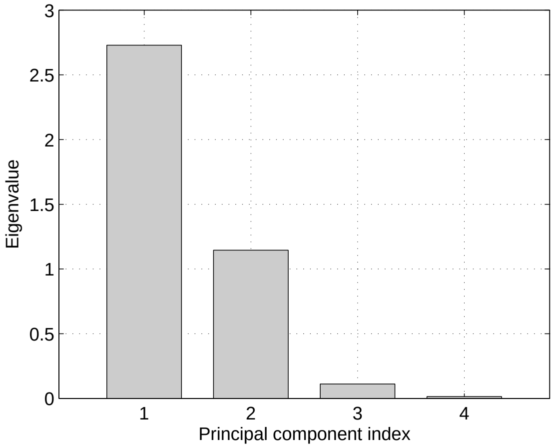

Figure 8: Eigenvalues of the four principal components.

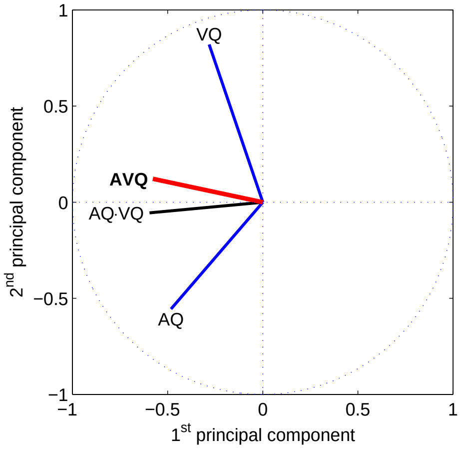

Figure 9: AQ, VQ, AQ·VQ as well as AVQ vectors relative to the first two principal components.

4.3.2. Modeling

The PCA described above provides evidence that both AQ and VQ contribute to AVQ. In this section, we further investigate this relationship in terms of modeling and prediction. As other researchers have proposed, $\mathrm{MOS_{AV}}$ can be modeled using $\mathrm{MOS_A}$, $\mathrm{MOS_V}$, and a multiplicative interaction term:[12, 15]

$$\widehat{\mathrm{MOS}}_{\mathrm{AV}} = a_0 + a_1\ \mathrm{MOS_A} + a_2\ \mathrm{MOS_V} + a_3\ \mathrm{MOS_A} \cdot \mathrm{MOS_V}. \tag{1}$$

We apply this model with different numbers of free parameters a_k to our data (a_0 is irrelevant for correlations, but improves the fit in terms of residual). The model accuracy of the various fits is shown in Figure 10. As expected from the results of the above PCA, good modeling is possible with only the multiplicative term:

$$\widehat{\mathrm{MOS}}_{\mathrm{AV}} = 1.98 + 0.103\ \mathrm{MOS_A} \cdot \mathrm{MOS_V} \tag{2}$$

or an additive linear model:

$$\widehat{\mathrm{MOS}}_{\mathrm{AV}} = -1.51 + 0.456\ \mathrm{MOS_A} + 0.770\ \mathrm{MOS_V}. \tag{3}$$

The latter provides a somewhat better fit, which is characterized by a correlation of 94% and an RMS residual of 0.44. The plane represented by Eqn. (3) is shown together with the actual $\mathrm{MOS_{AV}}$ values in Figure 11. It illustrates very well the higher importance attributed to VQ as compared to AQ. There is no improvement when using all four parameters in the fit.

We can compare these fits with other subjective experiments. Much of the existing work has focused on video-conferencing applications (i.e. head-and-shoulders clips), speech and/or simulated artifacts; the test material used here is quite different in terms of content range and distortions. Despite these significant differences, the multiplicative model from Eqn. (2) is very similar to previous models[12, 26] in terms of its parameters and goodness of fit. The same can be said about the additive model from Eqn. (3); the higher weighting of $\mathrm{MOS_V}$ over $\mathrm{MOS_A}$ is confirmed by other studies.[12, 15, 26] *

---

* Hands[15] shows that the content does have an influence on the model parameters, as he finds a stronger weighting of the audio component for video-conferencing material.

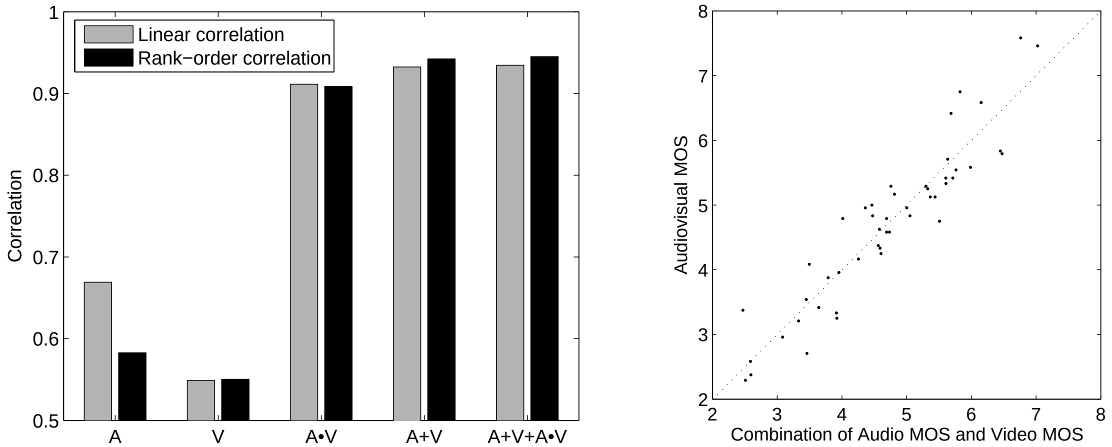

Figure 10: Correlations of different models for AVQ (left). A: model with MOS_A only ($a_2 = a_3 = 0$); V: model with MOS_V only ($a_1 = a_3 = 0$); A·V: multiplicative model from Eqn. (2); A+V: additive model from Eqn. (3); A+V+A·V: full model as in Eqn. (1) with all four parameters. The scatter plot of the additive model is shown on the right.

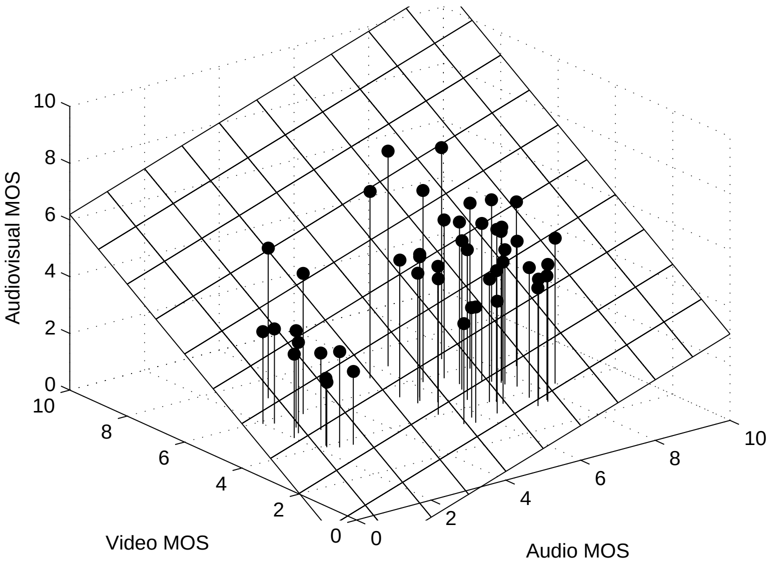

Figure 11: Plane of the AVQ model defined by Eqn. (3) and MOS_AV values (dots).

5. MOS PREDICTION

We now use the data obtained in our experiments to evaluate the MOS predictions of Genista's video and audio quality metrics.[*] In particular, we are interested in the prediction performance of our non-reference (absolute) metrics. The MOS predictions for video quality are based on a combination of artifact metrics such as jerkiness, blockiness[27] and blur[28] artifacts. These video metrics have been evaluated with several other datasets, including low-bitrate video.[24, 25] Similarly, the MOS predictions for audio quality are based on a combination of audio metrics such as bandwidth, loudness and saturation. All of these artifact metrics were designed to be computationally light, which makes it possible to compute them in real-time on a standard PC, while decoding and playing the clips.

Tuning is performed on a randomly selected half of the data, and the other half is used for evaluation. The resulting prediction performance is illustrated in Figure 12. First the video and audio metrics are considered

---

[*] See http://www.genista.com/ for more information.

separately. The predictions from the video metrics ($\mathrm{MOS}_\mathrm{V}^\mathrm{P}$) achieve correlations of above 90% with MOS_V; the audio metrics ($\mathrm{MOS}_\mathrm{A}^P$) reach 95%. When $\mathrm{MOS}_\mathrm{A}^P$ and $\mathrm{MOS}_\mathrm{V}^P$ are combined according to the models for audiovisual MOS found in the previous section, MOS_AV can be predicted with good accuracy. The best AVQ predictions with correlations of 90% are obtained directly from audio and video metrics together.

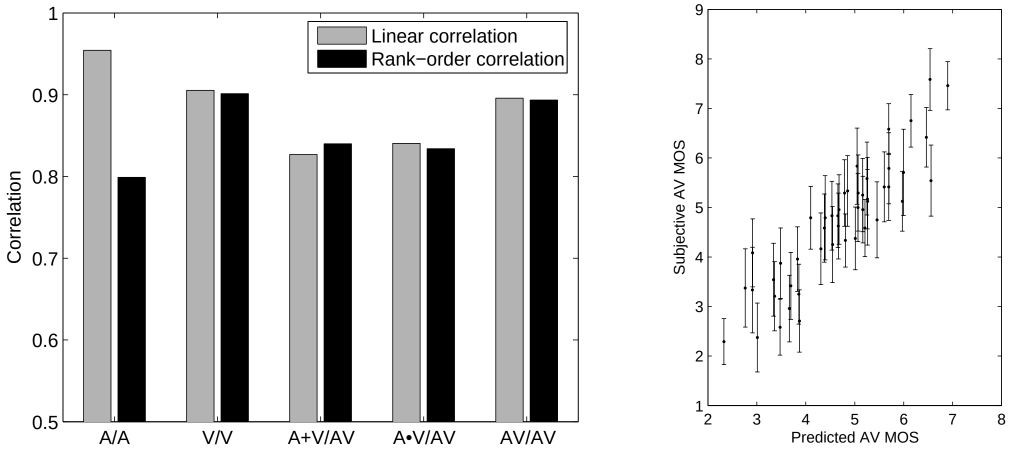

Figure 12: Prediction performance of Genista's non-reference metrics for video and audio. A/A: MOS_A predicted from audio metrics; V/V: MOS_V predicted from video metrics; A+V/AV, A·V/AV: MOS_AV predicted from $\mathrm{MOS}_\mathrm{A}^P$ and $\mathrm{MOS}_\mathrm{V}^P$ when combined according to Eqns. (3) and (2), respectively; AV/AV: MOS_AV predicted directly from audio and video metrics. The scatter plot corresponding to the last case is shown on the right.

6. CONCLUSIONS

We carried out subjective experiments on audio, video and audiovisual quality using the ACR methodology. Our main interest was mobile video transmission and the 3GPP format. We focused on MPEG-4 AVC/H.264 and MPEG-4 AAC to encode our test material at very low bitrates (24–48 kb/s for video and 8–32 kb/s for audio).

We found that both audio and video quality (AQ and VQ) contribute significantly to perceived audiovisual quality (AVQ). The product of AQ and VQ is an effective model of AVQ, and so is the linear combination of AQ and VQ. Our models confirm the results of previous studies, despite the substantial differences in test material.

Finally, we used non-reference metrics for audio and video to predict MOS. Their predictions match well with the subjective ratings for perceived audio, video, and audiovisual quality, achieving correlations of around 90%.

ACKNOWLEDGMENTS

We thank everyone who participated in our experiments. Prof. Sabine Süsstrunk at EPFL's Audiovisual Communications Lab provided some of the subjective testing facilities. Christophe Tournery resolved a bug in the testing software, and Remi Tyrand helped with the analysis of the audio quality data.

REFERENCES

1. S. Winkler, *Digital Video Quality – Vision Models and Metrics*, John Wiley & Sons, 2005.
2. VQEG, "Final report from the Video Quality Experts Group on the validation of objective models of video quality assessment," April 2000. Available at http://www.vqeg.org/.
3. VQEG, "Final report from the Video Quality Experts Group on the validation of objective models of video quality assessment – Phase II," Aug. 2003. Available at http://www.vqeg.org/.
4. ITU-R Recommendation BT.500-11, "Methodology for the subjective assessment of the quality of television pictures." International Telecommunication Union, Geneva, Switzerland, 2002.

5. ITU-T Recommendation P.910, "Subjective video quality assessment methods for multimedia applications." International Telecommunication Union, Geneva, Switzerland, 1999.
6. ITU-R Recommendation BT.1683, "Objective perceptual video quality measurement techniques for standard definition digital broadcast television in the presence of a full reference." International Telecommunication Union, Geneva, Switzerland, 2004.
7. ITU-T Recommendation J.144, "Objective perceptual video quality measurement techniques for digital cable television in the presence of a full reference." International Telecommunication Union, Geneva, Switzerland, 2004.
8. ITU-R Recommendation BS.1284-1, "General methods for the subjective assessment of sound quality." International Telecommunication Union, Geneva, Switzerland, 2003.
9. ITU-R Recommendation BS.1534-1, "Method for the subjective assessment of intermediate quality level of coding systems." International Telecommunication Union, Geneva, Switzerland, 2003.
10. ITU-T Recommendation P.862, "Perceptual evaluation of speech quality (PESQ), an objective method for end-to-end speech quality assessment of narrowband telephone networks and speech codecs." International Telecommunication Union, Geneva, Switzerland, 2001.
11. ITU-R Recommendation BS.1387-1, "Method for objective measurements of perceived audio quality (PEAQ)." International Telecommunication Union, Geneva, Switzerland, 2001.
12. J. G. Beerends and F. E. de Caluwe, "The influence of video quality on perceived audio quality and vice versa," *J. Audio Eng. Soc.* **47**, pp. 355–362, May 1999.
13. A. Joly, N. Montard, and M. Buttin, "Audio-visual quality and interactions between television audio and video," in *Proc. ISSPA*, pp. 438–441, (Kuala Lumpur, Malaysia), Aug. 13–16, 2001.
14. C. Jones and D. J. Atkinson, "Development of opinion-based audiovisual quality models for desktop video-teleconferencing," in *Proc. International Workshop on Quality of Service*, pp. 196–203, (Napa, CA), May 18–20, 1998.
15. D. S. Hands, "A basic multimedia quality model," *IEEE Trans. Multimedia* **6**, pp. 806–816, Dec. 2004.
16. ISO/IEC 14496-10, "Coding of audio-visual objects – Part 10: Advanced video coding." International Organization for Standardization, Geneva, Switzerland, 2004.
17. ITU-T Recommendation H.264, "Advanced video coding for generic audiovisual services." International Telecommunication Union, Geneva, Switzerland, 2003.
18. ISO/IEC 14496-2, "Coding of audio-visual objects – Part 2: Visual." International Organization for Standardization, Geneva, Switzerland, 2004.
19. ITU-T Recommendation H.263, "Video coding for low bit rate communication." International Telecommunication Union, Geneva, Switzerland, 1998.
20. ISO/IEC 14496-3, "Coding of audio-visual objects – Part 3: Audio." International Organization for Standardization, Geneva, Switzerland, 2001.
21. S. Winkler and C. Faller, "Maximizing audiovisual quality at low bitrates," in *Proc. Workshop on Video Processing and Quality Metrics*, (Scottsdale, AZ), Jan. 23–25, 2005.
22. ITU-T Recommendation P.911, "Subjective audiovisual quality assessment methods for multimedia applications." International Telecommunication Union, Geneva, Switzerland, 1998.
23. 3GPP Technical Specification 26.244, "Transparent end-to-end packet switched streaming service (PSS); 3GPP file format (3GP) (Release 6)." 3rd Generation Partnership Project, 2004.
24. S. Winkler and R. Campos, "Video quality evaluation for Internet streaming applications," in *Proc. SPIE*, **5007**, pp. 104–115, (Santa Clara, CA), Jan. 21–24, 2003.
25. S. Winkler and F. Dufaux, "Video quality evaluation for mobile applications," in *Proc. SPIE*, **5150**, pp. 593–603, (Lugano, Switzerland), July 8–11, 2003.
26. R. Pastrana-Vidal, C. Colomes, J. Gicquel, and H. Cherifi, "Caractérisation perceptuelle des interactions audiovisuelles: Revue," in *Proc. CORESA Workshop*, (Lyon, France), Jan. 16–17, 2003.
27. S. Winkler, A. Sharma, and D. McNally, "Perceptual video quality and blockiness metrics for multimedia streaming applications," in *Proc. International Symposium on Wireless Personal Multimedia Communications*, pp. 547–552, (Aalborg, Denmark), Sept. 9–12, 2001.
28. P. Marziliano, F. Dufaux, S. Winkler, and T. Ebrahimi, "A no-reference perceptual blur metric," in *Proc. ICIP*, **3**, pp. 57–60, (Rochester, NY), Sept. 22–25, 2002.