On the Feasibility of Video Streaming through Underwater Acoustic Links

Filippo Campagnaro, Roberto Francescon, Davide Tronchin, Michele Zorzi

***Abstract*—Nowadays video IP traffic accounts for 73% of the entire global Internet traffic with forecasts up to 82% by 2021 [1]. While live video streaming is already mature through media such as coaxial cables, optical fibers and radio links, real-time live video streaming through underwater acoustic communication is still in its infancy. The underdevelopment of underwater live streaming is due to both the unstable nature and the long propagation delay of the acoustic channel. The former poses several obstacles to reaching the needed bitrate capabilities, while the latter causes a non-negligible video latency proportional to the distance between transmitter and receiver. Despite these obstacles, a lot of research on advanced video codecs is conducted to reduce the required bitrates of a video stream. In addition, modem manufactures recently developed short range high rate acoustic modems.**

This work presents a feasibility study of a live video streaming based on the best performing video codecs (H.264/MPEG-AVC, H.265/MPEG-HEVC and VP9) through current commercial acoustic modems. The feasibility evaluation has been proved with a final pool test, where the video has been successfully streamed with real acoustic modems.

***Index Terms*—Underwater acoustic communications, video streaming, throughput, compression, EvoLogics, tank test.**

I. Introduction and related works

Underwater wireless communications have witnessed a wide and important development phase in the last decade, both at the physical [2] and at the upper layers of the ISO/OSI stack [3], as demonstrated by the introduction of the first underwater communication standard [4].

The acoustic modem improvements and developments have been pushed by a diversified set of application scenarios and led to a wide range of devices. In fact, even among the commercial off-the-shelf (COTS) acoustic modems, it is possible to find both very long-range and low-rate devices, and high-rate and short-range ones. For instance, the Develogic HAM.NODE [5] is able to transmit up to 30 km at 100 bit/s, while the Benthos ATM [6] modem provides a maximum bitrate of 15 kbit/s up to 1.5 km. At the same time, video streaming protocols have recently undergone a rapid development phase, due to the need to efficiently deliver media content to devices with different throughput capabilities such as smartphones and cars. The aim was to reduce the minimum required bitrate while, at the same time, maintaining an acceptable video quality.

Combination of both the recent development on short range COTS and the newly designed video streaming protocols, led to the need to investigate whether these technologies enable new applications such as video streaming through an acoustic underwater link.

BaltRobotics streamed a low-quality video with their acoustic prototype [7], able to achieve 115 kbit/s within a range of 200 m. The authors employed a customized self-developed video codec. MIT used a similar approach a few years before, developing a prototype able to perform video streaming through acoustics in a controlled scenario up to 200 m [8]. Northeastern University estimates that their SEANet modem prototype [9] will reach a data rate of 250 kbit/s at a range of 200 m, therefore supporting video streaming as well. So far they were able to transmit at 104 kbit/s with a bit error rate (BER) of 10^{-5}, while at 250 kbit/s the BER was 10^{-2} [2]. Further improvements are under development. In [10] the authors presented an acoustic modem prototype able to transmit up to 1 Mbit/s up to 100 m. They successfully transmit live video at 5 m distances in a wave tank in the presence of a high Doppler effect. In the same way, in [11] the authors proved their modem can stream high quality video in a 12 m pool. In [12] TNO performed a live data streaming by employing standard codecs through their medium rate acoustic modems (8 kbit/s). However, they managed to transfer only low a motion quasi-static video, that requires a low bitrate. EvoLogics GmbH developed the S2CM HS modem [13], which achieves 62.5 kbit/s up to 300 m and seems to be the best performing COTS modem available so far for short acoustic range communication.

Although recently other communication techniques have been developed, using optical [14], magneto-inductive [15] and electromagnetic (EM) [16] fields, in some scenarios acoustics still remains the only technology able to perform underwater wireless communications: for this reason, in this paper we present a feasibility study of a live video streaming through acoustics.

In the last years, new video codecs have been released. Among these, H.265 [17] and VP9 [18] are the mostly widely used and aim to replace H.264 [19] and VP8 [20], respectively. A quick overview of such codecs is presented in Section II. In Section III, both the test setup and the system configuration are reported, while the test results are presented in Section IV. Finally, in Section V, we draw our concluding remarks.

II. Video codecs overview

All modern video streaming systems employ some kind of source coding to reduce the minimum amount of information needed to visually represent the video. The algorithm used to perform this coding is composed by an *encoder*, at the

F. Campagnaro (Corresponding author, email: campagn1@dei.unipd.it), D. Tronchin and M. Zorzi are with the Department of Information Engineering, University of Padova, Italy. R. Francescon is with Wireless and More srl, Padova, Italy.

TABLE I
VIDEO SIZES FOR DIFFERENT HD ENCODINGS AND TWO DIFFERENT CHROMA SUBSAMPLING SCHEMES (UNCOMPRESSED 10-BITS SAMPLES) [23].

Frame size / frame rate	RGB(4:4:4) [Mbit/s]	YUV(4:4:2) [Mbit/s]
1280×720 / 60 p	211	141
1920×1080 / 24 PsF	190	127
1920×1080 / 50 i	198	132
1920×1080 / 60 i	237	158

streamer, and a *decoder*, at the consumer: the combination of the two forms a *codec*.

To reduce the amount of information transmitted it is necessary to remove the redundancy found in the original data; this redundancy can be found at two levels: among the pixels of the same frame and among pixels of different frames. The former is called *spatial redundancy* and is extensively used in image compression; the latter is called *time redundancy* and its exploitation is video-specific and calls for new dedicated algorithms.

The selected video codecs are the latest in terms of capabilities, namely: H.264/AVC, H.265/HEVC and VP9. H.264 [19] and H.265 [17] have been developed by the MPEG group with the stated intent of reducing the needed supporting bitrate, compared to previous standards. The VP9 codec, successor of VP8, has been developed by Google in response to the patent threat posed by the H.26x codec family [18].

The number of bits needed to represent a video signal depends on many factors, and can change significantly while maintaining the same perceived quality. Among more obvious parameters, such as the size of the frame or the frame rate, there are also more technical aspects, like the possibility to choose between different color encoding systems, such as the RGB color representation scheme [21] and the YUV chroma subsampling scheme [22]. The former is based on the trichromats human perception of colors, while the latter allows to reduce the information related to colors, which has less visual impact, to allow for more luminance data. Table I shows the bitrates for some common configurations: from these video settings, the YUV 4:4:2 scheme significantly reduces the needed bitrate, compared to a more basic RGB.

A. Analyzed parameters

Although *subjective parameters* are still widely used to evaluate video streaming performance, *objective parameters*, which include more classical measures such as SNR, will be used in the experiment evaluation phase: in particular, the Peak Signal-to-Noise Ratio (PSNR) will be analyzed. The PSNR is an index of the quality of the compressed image with respect to the uncompressed one, and is based on the Mean Squared Error (MSE) between a decoded frame and the original one, and on the number of bits, n, used to represent a sample; it is expressed as:

$$\mathtt{PSNR} = 10 \cdot \log_{10} \frac{(2^n - 1)^2}{\text{MSE}} \quad (1)$$

In case of color images, the MSE is the mean square error also over the color components (e.g., Y, U and V). Other two parameters analyzed in the evaluation phase are the average video bitrate (AvR), calculated as the video size, in bits, divided by the video duration, in seconds, the maximum instantaneous video bitrate (MaxR), the video quality metric (VQM [24]), an objective index of the video distortion (the bigger the index the higer the distortion), and the structural similarity index measure (SSIM [25]), a factor that shows the similarity of the two images, with acceptable values ranging from 0 (maximum difference) to 1 (no difference).

B. Video Description

The video employed in our test has been captured with a subsea camera during a remotely operated vehicle (ROV) operation. At second 27 the ROV resurfaced to perform a set of tasks above the water. The video size is 144 MB, with a duration of 70 s, and a resolution of 1920×1080 px at 29.97 fps. No audio was recorded, and the video uses a WMV3 codec with YUV 4:2:0. In this video AvR = 16.5 Mbit/s, and MaxR = 24 Mbit/s.

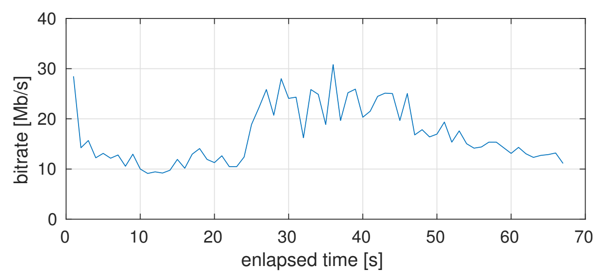

Fig. 1. Received bitrate during the raw video stream.

The instantaneous bitrate is not constant during the duration, as can be noticed from Figure 1. This behavior is related to the change of contents in the video itself, as a static video with low sporadic motions can be coded with a lower number of bits than a dynamic video with lots of movements. For this reason, in the same video, a static part with few details requires a smaller bitrate than a part of the video with fast motions and complex details. Indeed, the peak of the bitrate is related to the moment when the ROV resurfaced, and the video changed from a subsea video to a regular in-air video.

To measure these quantities, ITU-T P.910 defines two metrics: spatial perceptual information (SI) and temporal perceptual information (TI), both described in [26]. These two parameters, represented in Figure 2, play a crucial role in determining the amount of video compression that is possible, and consequently, the level of impairment that is suffered when the scene is transmitted over a fixed-rate transmission channel.

III. TESTS SETUP AND CONFIGURATION

The system evaluation has been realized in two steps. In the first step, a live video streaming has been performed between two PCs connected to the same Local Area Network (LAN). One PC, used as video transmitter, was directly connected to the LAN switch through an Ethernet cable, while the second PC, used as video receiver, was connected to the same network via WiFi, as presented in Figure 3.

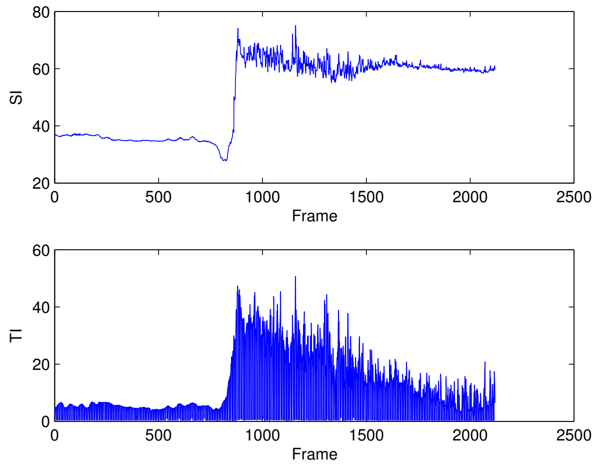

Fig. 2. Spatial perceptual information (SI) and temporal perceptual information (TI) of the video chosen for the test.

This first configuration has been employed to perform several video codec comparisons, in order to select the best candidate for the streaming through acoustics. This configuration has also been used to find the codec configuration that respects the bitrate target of an acoustic modem. The video has been streamed using UDP. Both results are presented in Section IV-A.

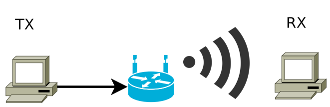

Fig. 3. LAN test setup.

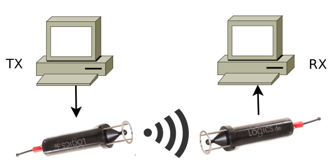

Fig. 4. Pool test setup.

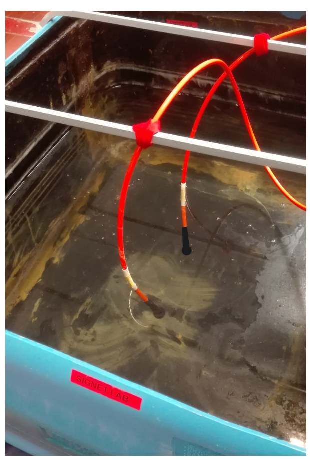

Fig. 5. Pool test picture.

The second evaluation test, described in Figure 4, has been performed using two EvoLogics S2CM HS underwater acoustic modems [13] submerged in a 60 cm×60 cm tank. This tank is internally covered with phono-absorbing material [27]; a picture of the test setup is presented in Figure 5.

The hardware employed in the tests included two EvoLogics HS acoustic modems, two PCs running the video streaming software, the small phono-absorbing tank and an Ethernet switch. Both PCs were provided with an 8 GB RAM, Intel core i7-4770k, Nvidia gtx 660 and Linux Mint. The software employed to perform the tests is listed in the following.

- VideoLAN (VLC) [28], open source multimedia player, encoder, and streamer, supporting many audio and video codecs, formats and streaming protocols. In our experiments, VLC used the FFmpeg [29] libavcodec library.
- Pipe Viewer (pv) [30], allows users to monitor the progress of file data reception, by giving information such as time elapsed, percentage completed, current throughput rate, total data transferred, and estimated time of arrival.
- netcat (nc) [31] and socat [32], utilities used to read and write data across network connections, using either TCP or UDP as the transport protocol.

IV. Results

A. Preliminary results and codec choice

The goal of this first evaluation is to select the best video codec candidate for an underwater acoustic live stream, in terms of coding efficiency, among H.264, VP9 and H.265. In this test VLC has been employed at both transmitter and receiver sides. In a first codec comparison, we streamed a video with 256×144 pixels resolution and 10 fps, setting the codec bitrate to 30 kbit/s, obtaining the results presented in Table II. With this configuration, the resulting video streaming is not real-time, as the receiver starts to reproduce the video 10 s after the beginning of the reception, due to the codec settings.

TABLE II
Video metrics for different codecs. Video resolution of 256×144 px, 10 fps.

Codec	AvR [kbit/s]	MaxR [kbit/s]	PSNR [dB]	SSIM	VQM
H.264	55.46	244	23.76	0.785	4.04
H.265	62.17	348	26.77	0.853	2.87
VP9	37.94	87	15.73	0.635	7.27

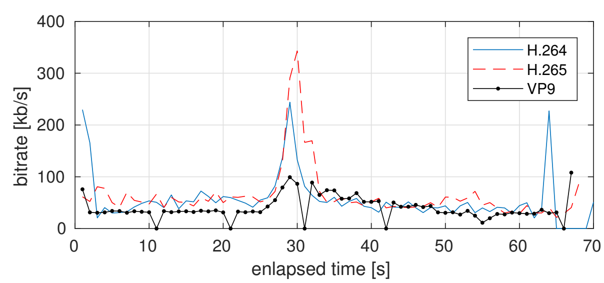

Fig. 6. Received bitrate during the video stream, with H.264 (solid blue line), H.265 (red dashed line) and VP9 (black line with point marks) codecs.

MaxR is reached in correspondence of the frames with the highest SI and TI, presented in Figure 2. This comparison highlights that H.265 provides a higher PSNR at the price of an AvR 2 times higher than the desired codec bitrate, set to 30 kbit/s. On the other hand, with the same video resolution, H.264 and VP9 provide a more efficient compression at the price of a lower video quality, due to the lower PSNR, the lower SSIM and a higher VQM. In particular, although in this configuration VP9 provides the lowest video quality, it provides the best compression and almost reaches the target bitrate. A possible reason why H.265 is outperformed by both H.264 and VP9 is that H.265 is more optimized for high definition videos than for low resolution video coding. Another reason might be that the implementation of H.265 for VLC is not yet as efficient as expected. In Figure 6, we compare the received bitrate during the streaming of the three codecs, where VP9 (represented with the solid black line with dot markers) is demonstrated to be more efficient than both H.264 (solid blue line) and H.265 (dashed red line).

To prove this assumption another test has been performed, issuing VLC to transmit the same video coded once with VP9 and once with H.264. The video resolution was set to 256×144 px and 7 fps, with a desired bitrate of 16 kbit/s. The results reported in Table III confirm the trend presented in Table II. In this case, although H.264 provides an `AvR` 20% lower than VP9, the latter has a `MaxR` 2.5 times smaller than the one obtained with H.264. Therefore, from these results we can expect that, although VP9 has a higher `AvR` than H.264, the former would more easily prevent the buffering issue [33]. VP9 has been pushed even further, by setting the codec bitrate to 12 kbit/s and framerate to 5 fps. Such configuration with H.264 did not perform well, as VLC was not able to perform the encoding and stream the video with less than 7 fps with this codec. We coded H.264 with this setting using FFMpeg, however, the resulting video received by VLC was neither smooth nor intelligible.

TABLE III
VIDEO METRICS FOR H.264 AND VP9 CODECS. VIDEO RESOLUTION OF 256×144 PX.

Codec	AvR [kbit/s]	MaxR [kbit/s]	PSNR [dB]	SSIM	VQM
H.264 7 fps, 16 kbit/s	19.12	156	16.02	0.771	4.13
VP9 7 fps, 16 kbit/s	24.15	62	15.42	0.632	8.00
VP9 5 fps, 16 kbit/s	21.13	59	15.58	0.637	7.34
VP9 5 fps, 12 kbit/s	17.71	44	15.63	0.679	6.67

B. Tank test

In the second evaluation test, described in Section III, we attempted to perform an underwater video streaming, by employing two EvoLogics HS acoustic modems, submerged in a 60 cm×60 cm tank. We let the modem stream with the EvoLogics data mode [34], by adapting the transmission bitrate according to the channel conditions. In this experiment, the modem achieved a bitrate of 31.25 kbit/s, over a maximum of 62.5 kbit/s. Although 31.25 kbit/s is higher than the `AvR` obtained with all the coding configurations presented in Table III, the actual maximum throughput reached during the transmission was about 18 kbit/s, due to the packet coding and header overhead. Although VP9 with a resolution of 256×144 px and a framerate of 5 fps provided an average bitrate of 17.71 kbit/s (slightly lower than the maximum throughput reached with the modem), and the first 20 s of the video were streamed successfully, the stream blocked in correspondence of the maximum SI and TI, due to the high bitrate requirement, and VLC crashed accordingly.

Nevertheless, with this configuration we tried to perform the streaming of a video with a lower motion during a completely subsea operation. This second attempt succeeded, as slow motions corresponds to quasi-static videos, with a low bitrate requirements due to the high correlation between frames. However, in order to obtain more robust and general results we had to stream a video with higher motions and, therefore, a higher maximum bitrate. Therefore, we performed the streaming of the original video with a lower resolution, employing VP9 at 200×96 px and framerate 5 fps. In Table IV, we present the performance metrics of this video. Although in this case `MaxR` = 42 kbit/s is higher than the maximum throughput, the video did not block and VLC was able to smoothly reproduce the video with few frame losses. Indeed, 14 frames, over a total of 645, have been discarded to avoid video jitter or additional delay. The stream reception started 10 s after the beginning of the transmission, but succeeded for the whole video duration. This delay is introduced by the video encoder itself, and not by the acoustic transmission, as also the dry streaming tests performed in Section IV-A provided the same delay. The received video bitrate during the VP9 stream through acoustics presented in Figure 7 is always bounded by the maximum throughput achieved by the modems in this experiment (18 kbit/s).

TABLE IV
VIDEO METRICS FOR H.264 AND VP9 CODECS. VIDEO RESOLUTION OF 200×96 PX, 5FPS AND CODING BITRATE SET TO 12 KBIT/S.

Codec	AvR [kbit/s]	MaxR [kbit/s]	PSNR [dB]	SSIM	VQM
VP9	13.68	42	19.22	0.754	6.88

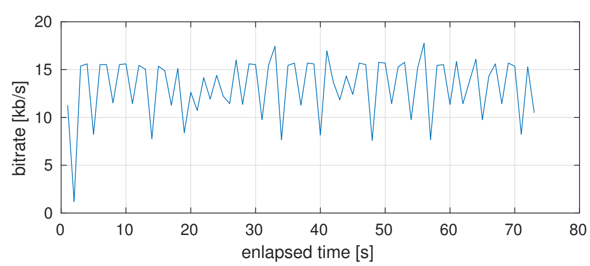

Fig. 7. Received bitrate during the low quality video stream via acoustics.

V. CONCLUSIONS AND FUTURE WORK

In this paper we discussed the feasibility of underwater acoustic live video streaming, in light of the capabilities offered by current COTS modems and the best performing standard video codecs. With each codec, we streamed the video with different settings, in order to define which configuration can be used to perform video streaming over an underwater acoustic channel. During a tank test, we were able to stream a quasi-real-time low quality video successfully.

Future tests will regard the further inspection of VP9 and H.265, testing them with different codec implementations and low latency configurations, to reduce the reception delay. In addition, the new standard AOMedia Video 1 (AV1) seems to be very promising [35], and it will be interesting to evaluate it in this scenario as soon as it will be released. In addition, performing the same test in the open sea or in a larger tank would help to obtain a better acoustic channel, and possibly stream the video at the maximum bitrate allowed by the modem.

REFERENCES

[1] "Cisco Visual Networking Index: Forecast and Methodology, 2016-2021," Last time accessed: June 2018. [Online]. Available: https://www.cisco.com/c/en/us/solutions/collateral/service-provider/visual-networking-index-vni/complete-white-paper-c11-481360.pdf

[2] E. Demirors, G. Sklivanitis, G. E. Santagati, T. Melodia, and S. N. Batalama, "A High-Rate Software-Defined Underwater Acoustic Modem with Real-Time Adaptation Capabilities," *IEEE Access*, vol. 6, no. 99, pp. 18 602–18 615, March 2018.

[3] R. Diamant, P. Casari, F. Campagnaro, and M. Zorzi, "Leveraging the nearfar effect for improved spatial-reuse scheduling in underwater acoustic networks," *IEEE Transactions on Wireless Communications*, vol. 16, no. 3, pp. 1480–1493, March 2017.

[4] J. Potter, J. Alves, D. Green, G. Zappa, I. Nissen, and K. McCoy, "The janus underwater communications standard," in *2014 Underwater Communications and Networking (UComms)*, Sept 2014, pp. 1–4.

[5] "Develogic Subsea Systems," Last time accessed: June 2018. [Online]. Available: http://www.develogic.de/

[6] "Teledyne-benthos acoustic modems," accessed: June 2018. [Online]. Available: https://teledynebenthos.com/product_dashboard/acoustic_modems

[7] "BaltRobotics," Last time accessed: June 2018. [Online]. Available: http://www.baltrobotics.com

[8] C. Pelekanakis, M. Stojanovic, and L. Freitag, "High rate acoustic link for underwater video transmission," in *Proceedings MTS/IEEE OCEANS.* IEEE, 2003.

[9] E. Demirors, B. G. Shankar, G. E. Santagati, and T. Melodia, "Seanet: A software-defined acoustic networking framework for reconfigurable underwater networking," in *Proc. ACM WUWNet*, Washington DC, US, Oct. 2015.

[10] J. Younce, A. Singer, T. Riedl, B. Landry, A. Bean, and T. Arikan, "Experimental results with HF underwater acoustic modem for high bandwidth applications," in *Proc. Asilomar Conf. on SS&C*, Pacific Grove, CA, Nov. 2015.

[11] M. Martins, J. Cabral, G. Lopes, and F. Ribeiro, "Underwater acoustic modem with streaming video capabilities," in *Proc. MTS/IEEE OCEANS*, Genova, Italy, May 2015.

[12] B. Binnerts, I. Mulders, K. Blom, M. Colin, and H. Dol, "Development and demonstration of a live data streaming capability using an underwater acoustic communication link," in *Proc. MTS/IEEE OCEANS*, Kobe, Japan, May 2018.

[13] "Evologics S2C M HS modem," accessed: June 2018. [Online]. Available: http://www.evologics.de/en/products/acoustics/s2cm_hs.html

[14] A. Caiti, E. Ciaramella, G. Conte, G. Cossu, D. Costa, S. Grechi, R. Nuti, D. Scaradozzi, and A. Sturniolo, "OptoCOMM: introducing a new optical underwater wireless communication modem," in *Proc. UComms*, Lerici, Italy, Sep. 2016.

[15] I. F. Akyildiz, P. Wang, and Z. Sun, "Realizing underwater communication through magnetic induction," *IEEE Communications Magazine*, vol. 53, no. 11, pp. 42–48, November 2015.

[16] X. Che, I. Wells, G. Dickers, P. Kear, and X. Gong, "Re-evaluation of RF electromagnetic communication in underwater sensor networks," *IEEE Communications Magazine*, vol. 48, no. 12, pp. 143–151, December 2010.

[17] G. J. Sullivan, J.-R. Ohm, W.-J. Han, and T. Wiegand, "Overview of the High Efficiency Video Coding (HEVC) Standard," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 22, no. 12, pp. 1649–1668, Sep. 2012.

[18] "WebM VP9 webpage," last time accessed: June 2018. [Online]. Available: https://www.webmproject.org/vp9/

[19] ITU-T Series H: Audiovisual and Multimedia Systems, *Advanced video coding for generic audiovisual services*, ITU Recommendation ITU-T H.264, Apr. 2017.

[20] "IBC2008: On2 Touts New Codec for Web Video," Last time accessed: June 2018. [Online]. Available: http://www.broadcastingcable.com/news/technology/ibc2008-on2-touts-new-codec-web-video/46250

[21] R. W. G. Hunt, *The Reproduction of Colour*. John Wiley & Sons Canada, Limited, Oct. 2004.

[22] "YUV pixel formats," Last time accessed: June 2018. [Online]. Available: http://www.fourcc.org/yuv.php

[23] "Uncompressed recorder," Last time accessed: June 2018. [Online]. Available: http://www.dvinfo.net/forum/blackmagic-design-hyperdeck-shuttle/494433-hyperdeck-shuttle-blackmagic-design-2.html

[24] M. Pinson and S. Wolf, "A New Standardized Method for Objectively Measuring Video Quality," *IEEE Transactions on Broadcasting*, vol. 50, no. 3, pp. 312–322, Sep. 2004.

[25] Z. Wang, L. Lu, and A. C. Bovic, "Video quality assessment using structural distortion measurement," in *International Conference on Image Processing*, 2002.

[26] "ITU-T P.910 - Subjective video quality assessment methods for multimedia applications," Last time accessed: June 2018. [Online]. Available: https://www.itu.int/rec/dologin_pub.asp?lang=e&id=T-REC-P.910-200804-I!!PDF-E&type=items

[27] F. Meneghello, F. Campagnaro, R. Diamant, P. Casari, and M. Zorzi, "Design and evaluation of a low-cost acoustic chamber for underwater networking experiments," in *Proc. ACM WUWNet*, Shanghai, China, Oct. 2016.

[28] "VLC media player," Last time accessed: June 2018. [Online]. Available: https://www.videolan.org/vlc/index.it.html

[29] "FFMpeg," Last time accessed: June 2018. [Online]. Available: https://www.ffmpeg.org/.

[30] "pv(1) - linux man page," Last time accessed: June 2018. [Online]. Available: https://linux.die.net/man/1/pv

[31] "The gnu netcat project," Last time accessed: June 2018. [Online]. Available: http://netcat.sourceforge.net/

[32] "socat(1) - linux man page," Last time accessed: June 2018. [Online]. Available: https://linux.die.net/man/1/socat

[33] "Definition of: buffering," accessed: June 2018. [Online]. Available: https://www.pcmag.com/encyclopedia/term/39024/buffering

[34] EvoLogics, "S2C reference manual," Jan. 2015, version 1.8.0.

[35] "Googles Royalty-Free Answer to HEVC: A Look at AV1 and the Future of Video Codecs," Last time accessed: June 2018. [Online]. Available: https://www.xda-developers.com/av1-future-video-codecs-google-hevc/