The Impact of Specialty Settings on the Perceived Quality of Medical Ultrasound Video

LUCIE LÉVÊQUE[1]**, (Student Member, IEEE), WEI ZHANG**[1]**, (Student Member, IEEE), PAMELA PARKER**[2]**, AND HANTAO LIU**[1]**, (Member, IEEE)**

[1]School of Computer Science and Informatics, Cardiff University, Cardiff CF24 3AA, U.K.
[2]Hull and East Yorkshire Hospitals NHS Trust, Hull Royal Infirmary, Hull HU3 2JZ, U.K.

Corresponding author: Lucie Lévêque (levequel@cardiff.ac.uk)

ABSTRACT Health care professionals are increasingly viewing medical images and videos in a variety of environments. The perception of medical visual information across all specialties, career stages, and practice settings are critical to patient care and patient safety. Visual signal distortions, such as various types of noise and artifacts arising in medical imaging, affect the perceptual quality of visual content and potentially impact diagnoses. To optimize clinical practice, it is of fundamental importance to understand the way medical experts perceive visual quality. Psychophysical studies have been undertaken to evaluate the impact of visual distortions on the perceived quality of medical images and videos. However, very little research has been conducted on how speciality settings affect the perception of visual quality. In this paper, we investigate whether and how radiologists and sonographers differently perceive the quality of compressed ultrasound videos, via a dedicated subjective experiment. The findings can be used to develop useful solutions for improved visual experience and better image-based diagnoses.

INDEX TERMS Medical image quality, subjective experiment, video quality assessment, ultrasound video, statistical analysis.

I. INTRODUCTION

Medical imaging is nowadays used in a broad range of medical specialties, including radiology, cardiology, pathology and ophthalmology [1]. In radiology, for example, there are approximately a billion imaging examinations conducted worldwide every year. The technologies used to acquire medical images include X-ray, ultrasound, computed tomography (CT), magnetic resonance imaging (MRI), positron-emission tomography (PET), single-photon emission computed tomography (SPECT), etc. Medical images provide an important source of information to assist clinicians with diagnostic decisions as well as recommendations for further action. However, they are not self-explanatory. Ultimately, medical images need to be inspected and interpreted by the human eye-brain system. Unfortunately, the interpretation task is not always easy and even competent clinicians make errors mainly due to the inherent limitations of human perception. Therefore, the decisions rendered by clinicians are not always absolutely conclusive [2]. To eliminate diagnostic errors and improve patient care, it is of fundamental importance to better understand perceptual factors underlying the creation and interpretation of medical images.

With the advent and growth of imaging technology in medicine, the methodologies used to acquire, process, transmit, store and display images vary and, consequently, the ultimate visual information received by clinicians or other health professionals differs significantly in perceived quality. Visual signal distortions, such as various types of noise and artifacts arising in medical image acquisition, processing, compression, transmission and rendering, affect the perceptual quality of images and potentially impact the accurate and efficient interpretation of images [3], [4]. Quality degradation of medical images often starts at the acquisition or image post-processing stage. For example, the common sources of MRI artifacts include non-ideal hardware characteristics, intrinsic tissue properties and their possible changes during scanning and a poor choice of scanning parameters [5], [6]. In digital radiology using X-ray, common artifacts are caused by e.g., under-exposure or over-exposure, collimation issues and grid use [7], [8]. In telemedicine, where medical images are being acquired, compressed, transferred and stored to diagnose and treat patients, various types of compression artifacts and transmission errors, such as blurring, ringing and packet loss, are produced [9].

In recent years, there has been a growing interest in studying the perceptual quality of medical images. Psychovisual experiments have been conducted with medical experts assessing the quality of images or videos in various application environments, e.g., in radiology or telemedicine [10]–[14]. The basic idea is to understand how the perception of medical image users is affected by specific visual distortions and then use what is learned to develop useful solutions for improved image quality and better image-based diagnosis. The study in [11] intends to find out the degree to which a medical image can be compressed, using JPEG or JPEG2000 algorithms, before its quality is compromised. A set of compressed CT images were presented to radiologists, who were requested to rate the quality of an image using a binary scale, i.e., acceptable or unacceptable. In the experiment, the radiologists were instructed to flag an image as unacceptable in the case they believed there was any noticeable distortion that could have any impact on diagnostic tasks. In [12], a series of image quality scoring experiments were performed to investigate the relative impact of different types of artifacts on the perceived quality of magnetic resonance (MR) images. MR images of different content affected with popular types of artifacts at different levels of energy were assessed by clinical application specialists. The study in [10] analyses the quality of real time wirelessly transmitted medical ultrasound video. Ultrasound trained medical professionals rated the quality of video, and the results were used to develop a minimum bit rate threshold to ensure transmitted video is of adequate quality so that physicians may make an accurate diagnosis. In [13], subjective image quality assessment was conducted with cardiologists. The perceptual effects of H.264 compression scheme [15] on echocardiographic and echo-Doppler sequences were investigated to identify a minimum bit rate that can preserve the diagnostic effectiveness of the ultrasound imaging sequences. In [14], medical ultrasound video sequences compressed via High Efficiency Video Coding (HEVC) scheme [16] were subjectively assessed by medical experts. The results were used to analyse the compression performance of HEVC in terms of acceptable diagnostic and perceptual video quality.

Obviously, the perception of medical image quality is critical to clinical practice. As health professionals are increasingly viewing medical images in a variety of environments, it is important to understand quality perception across speciality practice settings. Little progress has been made towards this purpose. In radiology practice, there are two groups of professionals who interact with and process image information. A radiologist is a doctor who is specially trained to interpret diagnostic images; and a radiographer is a person who has been trained to acquire medical images (note if a radiographer has been trained to perform an ultrasound, he/she may be called a sonographer). Both specialities are important for medical diagnosis. However, very little is known about the difference between radiologists and radiographers in term of their perception of image quality. In this paper, we investigate whether and to what extent specialty practice, i.e., radiologists versus sonographers, affect the quality perception of ultrasound video, through perception experimentation with compressed visual stimuli.

II. VISUAL QUALITY PERCEPTION EXPERIMENT: WITH BOTH RADIOLOGISTS AND SONOGRAPHERS
A. STIMULI

Unlike other related visual quality assessment studies in the literature which are either limited to a specific compression scheme or a small degree of stimulus variability, we aim to study a more comprehensive set of stimuli of a larger diversity in visual content and distortion. By this, we mean the dataset would include alternative popular compression schemes and various source stimuli and degradation levels. At the meantime, we seek to limit the total number of stimuli in order to make the subjective testing realistic so that the results are reliable. The source videos used in our experiments were extracted from four distinctive hepatic ultrasound scans by a senior radiologist from Angers University Hospital, Angers, France. To avoid potential bias, the radiologist would not be involved in the later stages of the experiments. It should be noted that, although the videos were from patients, they were purposely selected so that there was no apparent pathology. Also, the participants would not be informed of the indications for the scans. The reason behind above choices is to encourage the participants to consider all plausible clinical uses of the stimuli rather than focusing on a specific pathology. All source videos last twelve seconds each and have a resolution of 1920×1080 pixels at a frame rate of 25 frames per second (fps). Fig. 1 illustrates one representative frame of each source video.
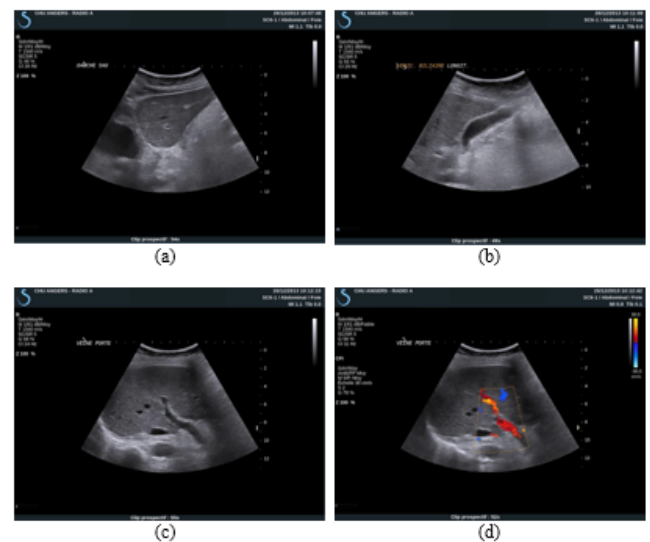

FIGURE 1. Illustration of one frame from each of the four source videos used in our experiment: (a) Content 1, (b) Content 2, (c) Content 3, and (d) Content 4 (in contrast to Content 1-3, Content 4 includes a Doppler ultrasound used to follow the blood flows).

The source videos were compressed using two popular compression schemes, namely H.264 [15] and HEVC [16].

H.264 is the most widely used video codec in current digital imaging systems, which allows for an efficient compression of visual signals due to its advanced functionalities in temporal and spatial predictions. Videos that are compressed by H.264 typically exhibit artifacts such as blocking, blur, ringing and motion compensation mismatches. The other video codec, HEVC, is the successor of H.264, and is meant to provide a better perceptual quality than H.264 at the same bit rate [17], [18]. HEVC compressed videos often exhibit mosquito noise around large regions of moving content. Both compression schemes could be potentially applied to the compression of clinical ultrasound video. To vary the perceptual video quality, for each source video, seven compressed sequences were created using the following bit rates: 512, 1000 and 1500 kbps (kilobits per second) for H.264 and 384, 512, 768 and 1000 kbps for HEVC. This resulted in a database of 32 video stimuli including the originals (i.e., 4 source videos $+4 \times 7$ compressed videos). It is well known that bit rate is not equal to quality for natural scenes, and that using the same bit rate to encode different natural contents could result in dramatically different visual quality. *However, studies on how compression can affect the quality perception of medical content, and to what extent that perception is dependent on the specific user group are largely unexplored.* The knowledge would be useful for the delivery of more usable visual content that is optimally rendered for the best performance and experience of clinical professionals.

B. EXPERIMENTAL PROCEDURE

Since standardised methodology for the subjective assessment of the quality of medical images and videos does not exist, we seek to make use of the experimental methodologies established for assessing natural images and videos. These methodologies are already described in detail in [19] and [20], where various experiment protocols are prescribed in order to suit different needs and environments of subjective visual testing while achieving consistent outcomes. The differences between diverse protocols include whether the reference (distortion free) stimulus is presented to participants when assessing the quality of the test (distorted) stimulus, and whether an absolute category rating scale or a continuous rating scale is used for scoring quality, etc. In making these choices and deciding on an appropriate protocol, factors that are often considered and traded off in practice include the ease of rating, timescale and reliability of data collection. To make our experiment feasible for radiologists, we conducted a user study where a few medical experts were surveyed for their preference in scoring quality of ultrasound videos. Based on the results of the survey, we decided to adopt a similar concept proposed by an established methodology, SAMVIQ (Subjective Assessment Methodology for Video Quality) [21], where video sequences are shown in multi-stimulus form. In contrast to other methodologies, the main advantage of SAMVIQ is that subjects can directly compare the related stimuli among themselves against the reference. Also, it allows the subjects to freely choose the order of

FIGURE 2. Illustration of the rating interface used in our experiment.

the tests and easily correct their votes as appropriate. These aspects are found relevant to the reading habits of experts in their clinical practice.

Fig. 2 illustrates the final scoring interface developed in our study. In the experiment, the subjects are asked to assess the overall quality of each video by inserting a slider mark on a vertical scale. The grading scale is continuous (with the score range [0, 100]) and is divided into three semantic portions to help clinical experts in placing their opinions on the numerical scale. The associated terms categorising the different portions are: ''Not annoying'' (i.e., [75, 100]) corresponding to ''the quality of the video enables you to conduct clinical practice without perceiving any visual artifacts''; ''Annoying but acceptable'' (i.e., [25, 75]) referring to ''the visual artifacts are noticeable but the quality of the video suffices for the conduct of clinical practice''; and ''Not acceptable'' (i.e., [0, 25]) meaning ''the visual artifacts are very noticeable and interfere with the clinical practice''. Fig. 2 also shows an example of the test organisation for each source scene, where an explicit reference (i.e., noted to the subjects), a hidden reference (i.e., a freestanding stimulus among other test stimuli) and seven compressed versions (placed in a different random order to each participant) are included. For each participant, the experiment is carried out scene after scene; and the order of scenes is randomised. Within a test (per scene), as shown in Fig. 2, subjects are allowed to view and grade any stimulus in any order; and each stimulus can be viewed and assessed as many times as the subject wishes (note the last score remains recorded). Note the entire methodology was developed in consultation with clinical experts to make sure the scoring experiment is more relevant and realistic to the reading environments in real clinical practice.

C. TEST ENVIRONMENT AND PARTICIPANTS

The experiment should be conducted in a typical radiology reading room environment. The venue should represent a controlled viewing environment to ensure consistent experimental conditions: low surface reflectance and approximately constant ambient light (i.e., with an indirect horizontal illumination of 100 lux). The stimuli were displayed on a Dell UltraSharp 27-inch wide-screen liquid-crystal display with a native resolution of 2560×1440 pixels, which
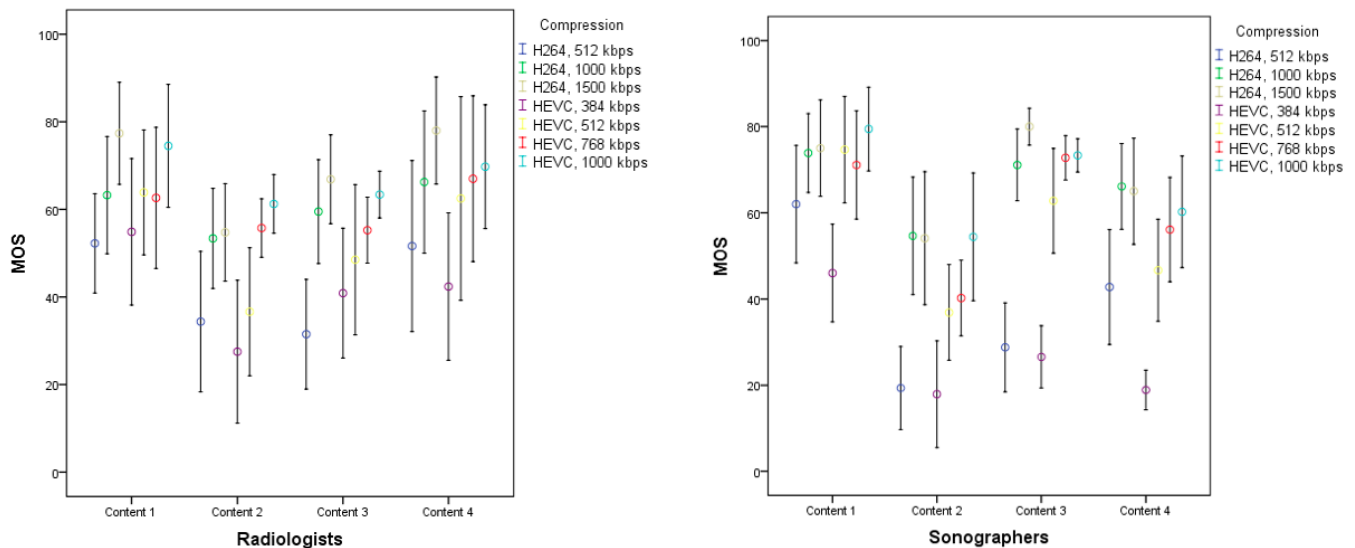
FIGURE 3. Illustration of the mean opinion score (MOS) averaged over all subjects (within a subject group, i.e., radiologists or sonographers) for each compressed video. "Content" refers to a source video. Error bars indicate a 95% confidence interval.

was calibrated to the Digital Imaging and Communications in Medicine (DICOM): Grayscale Standard Display Function (GSDF) [22]–[24]. The viewing distance was approximately 60 cm. No video adjustment (zoom, window level) was allowed.

Before the start of the actual experiment, each participant was provided with instructions on the procedure of the experiment (e.g., explaining the type of assessment and the scoring interface). A training session was conducted in order to familiarise the participants with the visual distortions involved and with how to use the range of the scoring scale. The video stimuli used in the training were different from those used in the real experiment. After training, all test stimuli were shown to each participant.

Since the goal of the study is to investigate visual quality perception across different specialities, our experiments were conducted with both radiologists and sonographers. Eight radiologists were recruited from Angers University Hospital, Angers, France, and nine sonographers from Castle Hill Hospital and Hull Royal Infirmary, Hull, United Kingdom. *Note that sample size used in our experiments is considered adequate in the area of medical image perception mainly due to the high degree of consistency among clinical experts [25], [26].*

III. IMAGE QUALITY ASSESSMENT BEHAVIOUR ANALYSIS: RADIOLOGISTS VERSUS SONOGRAPHERS

The two sets of raw data, one collected from radiologists and one from sonographers, were individually processed in the same way. First, a simple outlier detection and subject exclusion procedure was applied to the raw scores within a subject group [27], [28]. An individual score given for a video was considered an outlier if it was outside an interval of two standard deviations around the mean score for that

video. A subject was rejected if more than 20 percent of their scores were outliers. As a result of the outlier removal and subject exclusion procedure, none of the scores was detected as an outlier in both datasets and, therefore, no radiologist or sonographer was excluded from further analysis.

Fig. 3 illustrates the mean opinion score (MOS), averaged over all subjects (within a subject group), for each compressed video in our experiment. It can be seen clearly from Fig. 3 that sonographers appear to be more annoyed by the low-quality videos than radiologists, as sonographers scored the highly compressed videos (i.e., H.264: 512 kbps and HEVC: 384 kbps) lower in quality than radiologists. However, the difference is less obvious for the higher quality videos. The observed tendencies are further statistically analysed. In the case of the low-quality videos, i.e., H.264: 512 kbps and HEVC: 384 kbps, a statistical significance test is performed with the quality as the dependent variable and the specialty, i.e., radiologist vs. sonographer, as the dependent variable. As the test for the assumption of normality is not satisfied, a nonparametric version (i.e., the Mann-Whitney u-test) analogue to an independent samples t-test is conducted. The test results (i.e., statistic $= 2591$, p-value $= 0.004$) indicate that there is a statistically significant difference between radiologists and sonographers in rating low-quality videos. Similarly, in the case of higher quality videos, i.e., H.264: 1000 and 1500 kbps and HEVC: 512, 768 and 1000 kbps, preceded by a test for the assumption of normality, a Mann-Whitney u-test is performed and the results (i.e., statistic $= 13420$, p-value $= 0.207$) reveals that there is no statistically significant difference between radiologists and sonographers in rating higher quality videos.

Fig. 3 shows that compression settings – both variables of compression scheme and compression ratio – affect the video quality. Also, the effect tends to depend on video content,
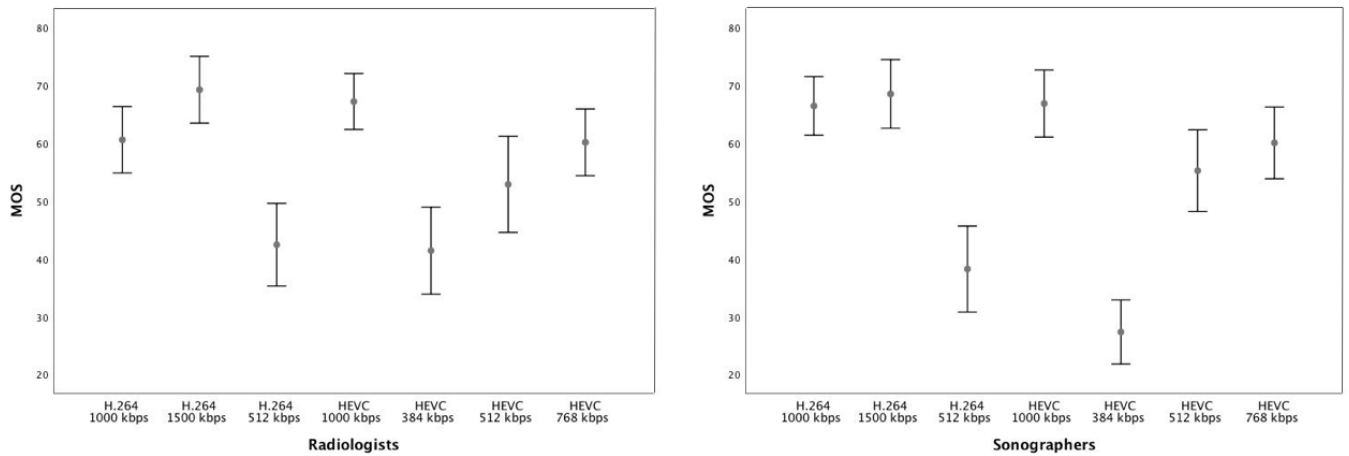
FIGURE 4. Illustration of the quality averaged over all subjects (within a subject group, i.e., radiologists or sonographers) and all contents for each compression configuration. Error bars indicate a 95% confidence interval.

TABLE 1. Results of the ANOVA to evaluate the effect of "Participant", "Content" and "Compression" on the perceived quality.

RADIOLOGISTS

Factor	df	F	p-value
Participant	7	1.365	0.266
Content	3	3.645	0.029
Compression	6	21.520	<0.01
Content * Compression	18	1.495	0.102

SONOGRAPHERS

Factor	df	F	p-value
Participant	8	1.264	0.306
Content	3	14.324	<0.001
Compression	6	68.959	<0.001
Content * Compression	18	3.862	<0.001

for example, in both cases of radiologists and sonographers, the quality of "Content 1" is consistently scored higher than the quality of "Content 2", independent of the compression scheme or compression ratio. Now, to further understand the impact of compression and content on video quality, we performed a statistical analysis, i.e., ANOVA (Analysis of Variance) using the software package SPSS version 23 [29]. In each case, the perceived quality is selected as the dependent variable, the video content and compression as fixed independent variables and the participant as random independent variable. The 2-way interactions of the fixed variables are included in the analysis. The results are summarised in Table I, where the F-statistic (i.e., F) and its associated degrees of freedom (i.e., df) and significance (i.e., p-value) are included.

First, in both cases, the results show that there is no statistically significant difference between participants in scoring

video quality (i.e., $p > 0.05$ in both cases). Note there is, therefore, little need to calibrate the raw scores using z-scores (as conventionally required for natural image and video quality assessment [30]) due to the consistency in scoring among individuals. This, in turn, reveals that the quality perception behaviour is highly consistent within a specialty group.

Second, the results show that all main effects (i.e., "Content" and "Compression" are statistically significant in each case. Not all source videos (i.e., "Content") have the same average quality (i.e., $p < 0.05$ in both cases). The post-hoc test reveals the following order in quality (note that commonly underlined entries are not significantly different from each other).

For radiologists: Content 2 ($\langle MOS \rangle = 46.23$) < Content 3 ($\langle MOS \rangle = 52.27$) < Content 4 ($\langle MOS \rangle = 62.5$) < Content 1 ($\langle MOS \rangle = 64.11$).

For sonographers: Content 2 ($\langle MOS \rangle = 39.65$) < Content 4 ($\langle MOS \rangle = 50.83$) < Content 3 ($\langle MOS \rangle = 59.33$) < Content 1 ($\langle MOS \rangle = 68.87$).

Clearly, both radiologists and sonographers score the quality of "Content 2" on average statistically significantly lower than the quality of other three source contents. "Content 1" tends to receive the highest quality scores in both cases. The impact of video content is probably due to the fact that different source videos may induce an intrinsic difference in sensitivity to distortion and thus in the annoyance of distortion. We can further observe a trend from the "unacceptable" quality scores (i.e., scores that are below 25) given by all participants that the majority of them are from one source video, i.e., eight scores and twelve scores from "Content 2", five scores and seven scores from "Content 3", three scores and six scores from "Content 4" and none from "Content 1" within the radiologists' and the sonographers' ratings, respectively. This implies that, using the same setting of video compression, "Content 2" is more likely to be affected by distortions. This perception is consistent between the two specialty groups.

TABLE 2. Results of statistical significance for pairwise comparisons (radiologists). Each entry in the table represents a code word consisting of three symbols: "1" means that the configuration for the row is statistically better than the configuration for the column, "0" means that it is statistically worse, and "-" means that it is statistically indistinguishable.

Compression	H264_512	H264_1000	H264_1500	HEVC_384	HEVC_512	HEVC_768	HEVC_1000
H264_512		0	0	-	0	0	0
H264_1000	1		0	1	1	-	0
H264_1500	1	1		1	1	1	-
HEVC_384	-	0	0		0	0	0
HEVC_512	1	0	0	1		0	0
HEVC_768	1	-	0	1	1		0
HEVC_1000	1	1	-	1	1	1	

TABLE 3. Results of statistical significance for pairwise comparisons (sonographers). Each entry in the table represents a code word consisting of three symbols: "1" means that the configuration for the row is statistically better than the configuration for the column, "0" means that it is statistically worse, and "-" means that it is statistically indistinguishable.

Compression	H264_512	H264_1000	H264_1500	HEVC_384	HEVC_512	HEVC_768	HEVC_1000
H264_512		0	0	1	0	0	0
H264_1000	1		-	1	1	1	-
H264_1500	1	-		1	1	1	-
HEVC_384	0	0	0		0	0	0
HEVC_512	1	0	0	1		-	0
HEVC_768	1	0	0	1	-		0
HEVC_1000	1	-	-	1	1	1	

Third, in either case, there is also a significant difference (i.e., $p < 0.05$ in both cases) in quality between the seven configurations of compression, and the post-hoc analysis reveals the following order in quality.

For radiologists: HEVC: 384 kbps ($\langle MOS \rangle = 41.41$) < H.264: 512 kbps ($\langle MOS \rangle = 42.44$) < HEVC: 512 kbps ($\langle MOS \rangle = 52.88$) < HEVC: 768 kbps ($\langle MOS \rangle = 60.16$) < H.264 1000 kbps ($\langle MOS \rangle = 60.60$) < HEVC: 1000 kbps ($\langle MOS \rangle = 67.22$) < H.264: 1500 ($\langle MOS \rangle = 69.25$).

For sonographers: HEVC: 384 kbps ($\langle MOS \rangle = 27.33$) < H.264: 512 kbps ($\langle MOS \rangle = 38.22$) < HEVC: 512 kbps ($\langle MOS \rangle = 55.25$) < HEVC: 768 kbps ($\langle MOS \rangle = 60.06$) < H.264: 1000 kbps ($\langle MOS \rangle = 66.44$) < HEVC: 1000 kbps ($\langle MOS \rangle = 66.86$) < H.264: 1500 ($\langle MOS \rangle = 68.53$).

The rankings of compression configurations (based on their average quality) tend to be highly consistent between radiologists and sonographers. Again, it is worth noticing here the difference in quality perception of low-quality videos between the two specialty groups. For HEVC: 384 kbps, sonographers score the quality on average much lower (i.e., $\langle MOS \rangle = 27.33$) than radiologists (i.e., $\langle MOS \rangle = 41.41$); similarly, for H.264: 512 kbps, sonographers score the quality on average lower (i.e., $\langle MOS \rangle = 38.22$) than radiologists (i.e., $\langle MOS \rangle = 42.44$). This indicates that radiologists show more tolerance of high distortions, whereas sonographers are more sensitive to highly distorted videos. At higher quality, sonographers are in close agreement with radiologists in terms of the average quality.

Finally, we investigate the impact of H.264 versus HEVC on the perceived quality of ultrasound videos. Fig. 4 illustrates the impact of the compression strategy on perceived quality, averaged over all subjects (within a subject group) and all source videos. For both cases, it can be seen that for each compression scheme (i.e., either H.264 or HEVC), the perceived quality monotonously increases with the increase of bit rate. Also, the following observations can be directly interpreted from Fig. 4. For radiologists, at low quality, one can conclude that the bit rate of H.264 (i.e., H.264: 512 kbps) should be 1.3 times as high as the bit rate of HEVC (i.e., HEVC: 384 kbps) to be perceived as equal quality. At high quality, to achieve the same perceived quality, the bit rate of H.264 (i.e., H.264: 1500 kbps) should be 1.5 times the bit rate of HEVC (i.e., HEVC: 1000 kbps). For sonographers, at high quality, to achieve the same perceived quality, the bit rate of H.264 (i.e., H.264: 1000 or 1500 kbps) should be 1 to 1.5 times the bit rate of HEVC (i.e., 1000 kbps). Pairwise comparisons are further performed with hypothesis testing between the two compression schemes, H.264 and HEVC. The results are summarised in Table II for the case of radiologists and Table III for the case of sonographers, where a paired samples t-test is performed if both samples are normally distributed; otherwise, in the case of non-normality, a nonparametric version analogue to a paired sample t-test (i.e., Wilcoxon signed rank sum) is conducted. For radiologists, Table II clearly indicates that there is no statistically significant difference in perceived quality between H.264: 512 kbps and HEVC: 384 kbps, and that similarly for the following pairwise comparisons: H.264: 1000 kbps

vs. HEVC: 768 kbps, and H.264: 1500 kbps vs. HEVC: 1000 kbps, the difference is not statistically significant for each case. For sonographers, Table III shows that there is no significant difference between H.264: 1000 kbps and H.264: 1500 kbps, and that similarly for the following pairwise comparisons: H.264: 1000 kbps vs. HEVC: 1000 kbps, H.264: 1500 kbps vs. HEVC: 1000 kbps, and HEVC: 512 kbps and HEVC 768 kbps, the difference is not statistically significant for each case.

IV. CONCLUSION

In this paper, we investigated how different medical specialty groups assess the quality of ultrasound video via a dedicated subjective experiment. We designed and conducted a perception experiment, where videos of different ultrasound exams distorted with various compression schemes and ratios were assessed by both radiologists and sonographers. For both specialty groups, the impact of visual content and compression configuration on the perceived quality of videos is found to be significant. Statistical analyses showed that the way the video quality changes with the content and compression configuration tends to be consistent for radiologists and sonographers. However, the results demonstrated that for the highly compressed (i.e., low quality) stimuli, sonographers are more annoyed by the distortions than the radiologists; and that for the moderately compressed (i.e., medium and high quality) stimuli, radiologists and sonographers behave similarly in terms of their quality of visual experience.

Our study provides new insights into the perception of medical video quality of health professionals, which can be used to optimise the experience of visual information in clinical practice. However, subjective visual testing is time-consuming and the evaluation is limited due to the amount and diversity of test stimuli. To facilitate further understanding of visual perception, future research should also focus on collecting and distributing more reliable subjective data.

REFERENCES

[1] E. A. Krupinski, "Current perspectives in medical image perception," *Attention, Perception Psychophys.*, vol. 72, no. 5, pp. 1205–1217, 2010.

[2] E. Samei and E. Krupinski, *The Handbook of Medical Image Perception and Techniques*. Cambridge, U.K.: Cambridge Univ. Press, 2010.

[3] E. Krupinski and Y. Jiang, "Anniversary paper: Evaluation of medical imaging systems," *Med. Phys.*, vol. 35, pp. 645–659, Feb. 2008.

[4] E. Krupinski, "Improving patient care through medical image perception research," *Policy Insights Behav. Brain Sci.*, vol. 2, pp. 74–80, Aug. 2015.

[5] E. Pusey *et al.*, "Magnetic resonance imaging artifacts: Mechanism and clinical significance," *Radiographics*, vol. 6, no. 5, pp. 891–911, 1986.

[6] J. Clark and W. Kelly, "Common artifacts encountered in magnetic resonance imaging," *Radiol. Clin. North Amer.*, vol. 26, pp. 893–920, Sep. 1988.

[7] S. L. Solomon, R. G. Jost, H. Glazer, S. S. Sagel, D. Anderson, and P. Molina, "Artifacts in computed radiography," *Amer. J. Roentgenol.*, vol. 157, no. 1, pp. 181–185, 1991.

[8] L. Cesar, B. Schueler, F. Zink, T. Daly, J. Taubel, and L. Jorgenson, "Artefacts found in computer radiography," *Brit. J. Radiol.*, vol. 74, pp. 195–202, Feb. 2001.

[9] L. Lévêque, W. Zhang, C. Cavaro-Ménard, P. Le Callet, and H. Liu, "Study of video quality assessment for telesurgery," *IEEE Access*, vol. 5, pp. 9990–9999, 2017.

[10] M. Gray, H. Morchel, and V. Hazelwood, "Evaluating the effect of bit rate on the quality of portable ultrasound video," in *Proc. IEEE 12th Int. Symp. Biomed. Imag.*, New York, NY, USA, Apr. 2015, pp. 247–250.

[11] I. Kowalik-Urbaniak *et al.*, "The quest for 'diagnostically lossless' medical image compression: A comparative study of objective quality metrics for compressed medical images," *Proc. SPIE*, vol. 9037, p. 9037L, 2014.

[12] H. Liu, J. Koonen, M. Fuderer, and I. Heynderickx, "The relative impact of ghosting and noise on the perceived quality of MR images," *IEEE Trans. Image Process.*, vol. 25, no. 7, pp. 3087–3098, Jul. 2016.

[13] A. Neri, F. Battisti, M. Carli, M. Salatino, M. Goffredo, and T. D'Alessio, "Perceptually lossless ultrasound video coding for telemedicine applications," in *Proc. Int. Workshop Video Process. Quality Metrics*, 2007.

[14] M. Razaak and M. Martini, "Rate-distortion and rate-quality performance analysis of HEVC compression of medical ultrasound videos," in *Proc. 4th Int. Conf. Sel. Topics Mobile Wireless Netw.*, Rome, Italy, 2014, pp. 230–236.

[15] T. Wiegand, G. J. Sullivan, G. Bjøntegaard, and A. Luthra, "Overview of the H.264/AVC video coding standard," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 13, no. 7, pp. 560–576, Jul. 2003.

[16] *High Efficiency Video Coding*, Document Rec. ITU-T H.265, 2013.

[17] G. J. Sullivan, J.-R. Ohm, W.-J. Han, and T. Wiegand, "Overview of the High Efficiency Video Coding (HEVC) standard," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 22, no. 12, pp. 1649–1668, Dec. 2012.

[18] J. Nightingale, Q. Wang, C. Grecos, and S. Goma, "The impact of network impairment on quality of experience (QoE) in H.265/HEVC video streaming," *IEEE Trans. Consum. Electron.*, vol. 60, no. 2, pp. 242–250, Feb. 2014.

[19] *Methodology for the Subjective Assessment of the Quality of Television Pictures*, Document Rec. ITU-R, BT.500-11, 2002.

[20] *Subjective Video Quality Assessment Methods for Multimedia Applications*, Document Rec. ITU-T 910, 2008.

[21] F. Kozamernik, V. Steinmann, P. Sunna, and E. Wyckens, "SAMVIQ—A new EBU methodology for video quality evaluations in multimedia," *Soc. Motion Picture Television Eng.*, vol. 114, pp. 152–160, Apr. 2005.

[22] E. Samei, "Assessment of display performance for medical imaging systems," *Rep. Amer. Assoc. Physicists Med., Task group*, vol. 18, 2005.

[23] B. M. Hemminger, R. E. Johnston, J. P. Rolland, and K. E. Muller, "Introduction to perceptual linearization of video display systems for medical image presentation," *J. Digit. Imag.*, vol. 8, no. 1, pp. 21–34, 1995.

[24] (2004). *National Electrical Manufacturers Association, Digital Imaging and Communications in Medicine (DICOM) Part 14: Greyscale Standard Display Function, Rosslyn, USA*. Accessed on Jun. 2017. [Online]. Available: http://medical.nema.org/dicom/2004/04_14PU.pdf

[25] A. Kumcu *et al.*, "Visual quality assessment of H.264/AVC compressed laparoscopic video," *Proc. SPIE*, vol. 9037, pp. 1–12, Mar. 2014.

[26] L. Platisa, L. Van Brantegem, A. Kumcu, R. Ducatelle, and W. Philips, "Influence of study design on digital pathology image quality evaluation: The need to define a clinical task," *J. Med. Imag.*, vol. 4, no. 2, p. 021108, 2017.

[27] Z. Wang and A. Bovik, *Modern Image Quality Assessment: Synthesis Lectures on Image, Video and Multimedia Processing*. Morgan & Claypool, 2006.

[28] H. R. Sheikh, M. F. Sabir, and A. C. Bovik, "A statistical evaluation of recent full reference image quality assessment algorithms," *IEEE Trans. Image Process.*, vol. 15, no. 11, pp. 3440–3451, Nov. 2006.

[29] *IBM SPSS Statistics for Windows, Version 23.0*, IBM Corp., Armonk, NY, USA, 2014.

[30] A. van Dijk, J. Martens, and A. Watson, "Quality assessment of coded images numerical category scaling," *Proc. SPIE*, vol. 2451, pp. 90–101, Mar. 1995.

LUCIE LÉVÊQUE (S'17) received the B.Eng. and M.Eng. degrees in cognitics from the Ecole Nationale Supérieure de Cognitique, Bordeaux, France, in 2013 and 2015, respectively, and the M.Sc. degree in biomedical imaging from the University of Angers, Angers, France, in 2015. She is currently pursuing the Ph.D. degree with the School of Computer Science and Engineering, Cardiff University, Cardiff, U.K. Her research interests include visual media quality assessment and medical image perception.
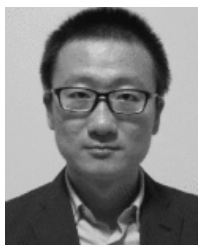
WEI ZHANG (S'14) received the B.Sc. and M.Sc. degrees from Xidian University, Xi'an, China, in 2011 and 2013, respectively, and the Ph.D. degree from Cardiff University, Cardiff, U.K, in 2017. His research interests include image analysis, video processing, and human visual perception.

PAMELA PARKER received the Diploma degree from the College of Radiographer in 1993 and the M.Sc. degree in medical ultrasound from the University of Leeds, Leeds, U.K., in 1999. She is currently a Clinical Sonographer and the Speciality Manager in ultrasound with the Hull and East Yorkshire Hospitals NHS Trust, Hull, U.K. She is currently a British Medical Ultrasound Society Professional Development Officer, the Chair of the Professional Standards Committee, and a member of the Editorial Board of *Ultrasound*.

HANTAO LIU (S'07–M'11) received the Ph.D. degree from the Delft University of Technology, Delft, The Netherlands, in 2011. He is currently an Assistant Professor with the School of Computer Science and Informatics, Cardiff University, Cardiff, U.K. He is currently serving the IEEE MMTC as the Chair of the Interest Group on Quality of Experience for Multimedia Communications and an Associate Editor of the IEEE Transactions on Human-Machine Systems and the IEEE Transactions on Multimedia. His research interests include visual media quality assessment, visual attention modeling and applications, visual scene understanding, and medical image perception.

● ● ●