Subjective Evaluation of HEVC in Mobile Devices

Ray Garcia, Hari Kalva
Florida Atlantic University, 777 Glades Road, Boca Raton, Florida, United States 33431

ABSTRACT

Mobile compute environments provide a unique set of user needs and expectations that designers must consider. With increased multimedia use in mobile environments, video encoding methods within the smart phone market segment are key factors that contribute to positive user experience. Currently available display resolutions and expected cellular bandwidth are major factors the designer must consider when determining which encoding methods should be supported. The desired goal is to maximize the consumer experience, reduce cost, and reduce time to market. This paper presents a comparative evaluation of the quality of user experience when HEVC and AVC/H.264 video coding standards were used. The goal of the study was to evaluate any improvements in user experience when using HEVC.

Subjective comparisons were made between H.264/AVC and HEVC encoding standards in accordance with Double-stimulus impairment scale (DSIS) as defined by ITU-R BT.500-13. Test environments are based on smart phone LCD resolutions and expected cellular bit rates, such as 200kbps and 400kbps.

Subjective feedback shows both encoding methods are adequate at 400kbps constant bit rate. However, a noticeable consumer experience gap was observed for 200 kbps. Significantly less H.264 subjective quality is noticed with video sequences that have multiple objects moving and no single point of visual attraction. Video sequences with single points of visual attraction or few moving objects tended to have higher H.264 subjective quality.

Keywords: HEVC, H.264, Mobile, Bit Rate, Subjective, Video Compression, TISI

1. INTRODUCTION

High Efficiency Video Coding (HEVC) is the next coding standard that is being finalized by the collaborative work between the International Organization for Standardization (ISO) and the International Telecommunication Union (ITU). This paper compares the H.264/AVC, a widely used coding standard for mobile video services, and HEVC in the smart phone market segment with a range of expected cellular constant bit rates, such as 200kbps and 400kbps. In addition, content characteristics of video sequences and their impact on the quality of user experience are categorized.

For this study, the focus within the mobile compute environment is smart phones. Multimedia use in smart phones has increased with the surge in popularity for these devices. A significant contributor is the dramatic improvement for LCD performance. Recent mobile devices released in the consumer market have shown display technology has progressed strongly. The displays range from high-end mobile phones with resolutions up to 1280x720 for 5.0" diagonal screen sizes to entry level smart phones with resolutions around 480 x 270 for 3.5". Also, cellular network bandwidth has evolved to support 3G (or better) for significant portion of heavily populated areas within the world. This leads to higher available bandwidth for use on mobile devices.

2. RELATED WORK

The Joint Collaborative Team on Video Coding (JCT-VC) reported subjective test results for 27 test candidates in April 2010 [1]. The purpose was to evaluate the candidates for the next generation encoder standard, HEVC. Two anchor encodings were generated to assist with the test candidate evaluations. Anchor encodings were included in the formal subjective tests and were directed through the same evaluation criteria as the test candidates. The H.264/AVC encoder used for anchor file generation is JM16.2. These anchor reference points were used to define behavior of current and accepted video encoding technologies for side-by-side comparison with test candidates.

Test methods used within test sessions were Double Stimulus Continuous Quality Scale (DSCQS) and Double Stimulus Impairment Scale (DSIS) evaluation methods as defined by ITU-R BT.500-13 [2]. DSIS test evaluation methods were used for class C (832 x 480 video sequence), D (416 x 240 video sequence), E (1280 x 720 video sequence), and lower bit rates for class B (1920 x 1080 video sequence). DSCQS test evaluation methods were used for higher encoding rates

Multimedia Content and Mobile Devices, edited by David Akopian, et al., Proc. of SPIE-IS&T Electronic Imaging,
SPIE Vol. 8667, 86670L · © 2013 SPIE-IS&T · CCC code: 0277-786X/13/$18 · doi: 10.1117/12.2003988

within class B sequences. Results from the evaluation showed a 50% bit rate improvement can be achieved with multiple test candidates and achieve similar mean opinion score (MOS).

Tan et al [3] objectively and subjectively measured performance and made comparisons between HM5.0 and JM18. Tests were conducted with high-efficiency (HE), low-complexity (LC) and low-complexity combinations of rate distortion optimized quantization (RDOQ), adaptive loop filter (ALF), sample adaptive offset (SAO). Coding tools were manipulated to improve BD-rate vs. encoding and decoding time. The study shows video encoded with HE configuration yielded subjectively indistinguishable results when compared to LC with RDOQ and SAO. Nine video sequences from class B and C were encoded with QP = 32 and 37 for both random access and low delay configurations. The target bit rates were predominantly 500 kbps to 4,000 kbps. The test method was DSIS variant I, which shows the reference and impaired video once to the test subject before voting takes place. Informal subjective tests used HM 5.0 encoded video sequences that are half the bit rate versus JM 18.2 encoded video sequences. Observer votes showed that HEVC is preferred over AVC/H.264 from 56% to 83% of the time. The preference percentage depends on the video sequence encoding configuration being used for the subjective test.

Additional subjective comparisons were performed by the JCT-VC ad hoc group comparing HM5 and similarly configured JM encoder/decoder and reported by Ohm, Sullivan et al [4]. The goal was to quantify feasible rate savings that yield similar subjective quality when comparing HEVC and similarly configured AVC/H.264. Tests were performed with the nine video sequences for class B and C. JM QP settings were 27, 30, 33, and 36. The research team determined JM QP settings should be four more. Therefore, $QP_{HM} = QP_{JM} + 4$, which gives a HM QP settings of 31, 34, 37, and 40. Subjective tests were performed using Double Stimulus Impairment Scale (DSIS) method with same approach described in [1]. After some linear interpolation on RD graphs, a gross average rate reduction of 67% for class B sequences and 49% for class C sequences were deduced.

Informal subjective tests for 720p and 1080p resolutions specifically targeting low-delay applications were performed by Horowitz, Kossentini et al [5]. H.264 / AVC JM version 18.3 and x264 version core 122 r2184 are compared with HEVC (HM version 7.1). Encoders were configured for low-delay function and 8-bit per sample video encoding. The AVC/H.264 videos were encoded at double the rate of the HEVC videos. QP was selected to ensure lossy video, but still considered good quality video. Very high quality video will yield experiment results where both video sequences are indistinguishably excellent video. Also, extremely low quality video was avoided. Poor quality video will be difficult for the observer to indicate a preference. Results indicate HM encoded sequences were favored 46% of the time for 720p sequences and 86% of the time for 1080p sequences.

Resolutions beyond HDTV are the main emphasis for subjective analysis by Hanhart et al [6]. Tests were conducted on a high performance quad full high definition (QFHD) liquid crystal display (LCD). Video sequences were evaluated for spatial information (SI) and temporal information (TI) indexes. Class A video sequences were used, since these sequences have the highest resolution. Also, test video sequences were augmented with two other high resolution sequences for a video resolution of 3840x1744 (or higher). Bit rates for tests were encoded and ranged from 768 kbps to 20Mbps. Double Stimulus Impairment Scale DSIS variant II, as defined by [2], was the test method and the test session was divided into two 15 minute sessions with a rest period in between. Test results convincingly showed reduction of over 50% is achieved with HEVC over AVC for high resolution sequences for equivalent subjective performance.

3. MOTIVATION

A significant amount of H.264/AVC and HEVC subjective testing has been conducted for video encoded at higher bit rates (500 kbps or higher) for resolutions of 832 x 480 or higher. The mobile multi-media environments study has been underserved and not been sufficiently examined to determine the impact of lower bit rate availability and lower resolution capability within the smart phone environment. Understanding the subjective test implications will permit effective use of mobile designer's choices.

4. METHODS

4.1 Background

Mobile encoding method recommendations, within this study, revolved around recommendations from multimedia streaming guidelines and content providers which define "higher quality" as 640x360 resolution at 400kbps and "medium quality" as 400x300 resolution at 200kbps [7] [8]. Smart phone display performance has progressed significantly in recent years which led the research team to choose the higher resolution (640 x 360) to encode the test video sequences. Wireless bandwidth providers will have challenges in providing the highest level of service with current and future wireless (mainly cellular) networks. This will lead to the desired preference of reducing streaming content bandwidth as much as possible while minimizing user subjective performance deficits. The expectation is for video sequences to have a wireless transmission bit rate in the realm of 200kbps and 400kbps. 200 kbps is considered a constrained transmission bandwidth and 400 kbps is considered a good transmission rate for mobile resolution encoded sequences. Both 200 and 400 kbps were chosen for subjective tests.

Encoder revisions used for the comparison are H.264/AVC Software Coordination version: JM18.3 [9] and HEVC version HM 6.0[10]. H.264/AVC was configured to closely emulate HEVC coding based on HM-like configurations available in JM 18.3. The benefit is to reduce configuration variability between the two video encoding protocols.

The video sequences chosen are from ITU test suite. A total of eight sequences are in the experiment pool. These are: Basketball Drill, Flowervase, Keiba, Kimono, Johnny, People on Street, Race Horses, and Traffic. Video sequence characteristics, shown in Table 1, include a frame per second (fps) range from 24 fps to 60 fps and were not altered from the source. HEVC video test sequences chosen were scaled and cropped to 640x360 resolutions. As mentioned earlier, this is chosen since 640 x 360 resolution is considered high quality resolution for the mobile market. Encoding to this resolution will factor significantly in data needed per video sequence, which will be a strong factor for transmission bit rate. Essentially, the larger the resolution, the more data needs to be transmitted per frame.

Table 1: Video sequence characteristics

Video Sequence	Width	Height	fps	Frames
Basketball Drill	640	360	50	500
Flowervase	640	360	30	300
Keiba	640	360	30	300
Kimono	640	360	24	240
Johnny	640	360	60	600
People On Street	640	360	30	150
Race Horses	640	360	30	300
Traffic	640	360	30	150

The video test sequences were encoded at various quality levels. The H.264 videos were encoded with QP varied from 27 to 51. For HEVC encoded video sequences, the QP range is 24-51. From this effort, the video sequences with bit rate closest to 200 kbps and 400 kbps were chosen and shown in Table 2 and Table 3. Also, care was taken to minimize the bit rate delta between the H.264 and HEVC video sequences in order to evaluate performance at target bitrates of 200 and 400 Kbps.

Table 2 : 400kbps data for H.264 and HEVC average bitrate, bitrate difference. Delta = HEVC(avg. bitrate) – H.264(avg. bitrate).

Test	PSNR(Y)	Avg. Bitrate	Delta
BBDrill-H264	29.497	380.45	21.26
BBDrill-HEVC	31.4938	401.71	
Flowervase-H.264	35.423	349.34	36.73
Flowervase-HEVC	37.6337	386.07	
Johnny-H264	39.036	398.72	2.17
Johnny-HEVC	39.4097	400.89	
Keiba-H.264	32.965	434.95	-13.54
Keiba-HEVC	34.3003	421.41	
Kimono-H264	32.737	408.93	-8.56
Kimono-HEVC	33.2376	400.37	
People-H.264	20.77	368.99	-15.81
People-HEVC	21.2125	353.18	

RaceHorses-H.264	28.6	390.22	-5.29
RaceHorses-HEVC	29.6388	384.93	
Traffic-H.264	27.256	387.70	-32.14
Traffic-HEVC	28.3356	355.56	

Table 3 : 200kbps data for H.264 and HEVC average bitrate, bitrate difference. Delta = HEVC(avg. bitrate) – H.264(avg. bitrate).

Test	PSNR(Y)	Avg. Bitrate	Delta
BBDrill-H264	26.887	198.78	2.80
BBDrill-HEVC	29.432	201.58	
Flowervase-H.264	33.323	204.73	-14.49
Flowervase-HEVC	35.1644	190.24	
Johnny-H264	36.614	193.37	7.70
Johnny-HEVC	37.7343	201.07	
Keiba-H.264	29.66	211.05	-5.93
Keiba-HEVC	31.5785	205.12	
Kimono-H264	30.429	198.41	1.29
Kimono-HEVC	30.961	199.70	
Mobisode2_H.264	40.982	180.84	3.18
Mobisode2-HEVC	42.5948	184.02	
People-H.264	19.15	203.14	-0.57
People-HEVC	19.7234	202.57	
RaceHorses-H.264	26.382	188.23	10.42
RaceHorses-HEVC	27.6485	198.65	
Traffic-H.264	25.458	193.08	-0.53
Traffic-HEVC	26.6407	192.55	

Observers were shown the impaired sequences in accordance with double-stimulus impairment scale (DSIS) Variant II as defined by ITU-R BT.500-13[2]. The double stimulus method is cyclic where the observer is presented with an unimpaired reference followed by the same sequence but impaired. Variant II was chosen to allow observer a second viewing of both video sequences. For tests within this study, a 3 second grey background is added between each video sequence. The presentation order of the video sequence is shown in Figure 1. Basically, the video sequence starts with 3 seconds of solid grey frames, followed by 5 or 10 seconds of the reference sequence, followed by 3 seconds of solid grey frames, followed by 5 or 10 seconds of the impaired sequence, the previous four viewing events are repeated, followed by the voting cycle.

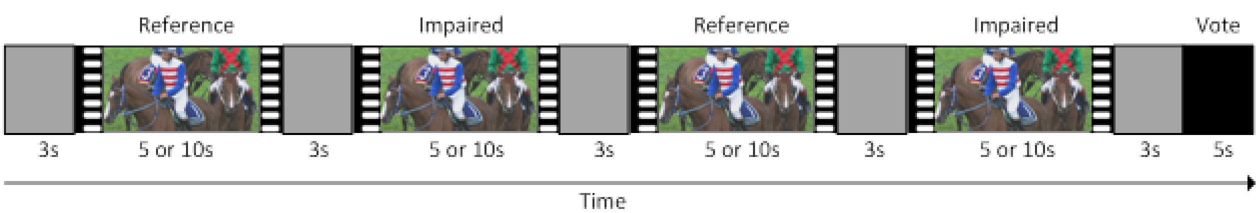

Figure 1: Test sequence presentation to observer

Observers (test subjects) were given a 2-3 minute explanation of the test structure and instructions for viewing and voting sections. Observers were allowed to vote after viewing the first sequence pair, which is the reference video sequence followed by the impaired video sequence. However, during the explanation session observers were encouraged to view the complete video set and vote during the voting period. The test sessions for observers were limited to 20 minutes. Several technical literature recommends a test session length to be 30 minutes or less. Test session length chosen is 2/3 of a 30 minute session to reduce chance of observer fatigue.

The impairment scale is a scale of 5 to 1, where the subjective differential steps are roughly the same. A rating of five indicates the reference video and impaired video were imperceptibly equivalent. By contrast, a rating of one indicates that the impaired video is very annoying when compared to the reference video. Voting feedback must select whole number ratings. Figure 2 shows the rating scale.

Score	Definition
5	imperceptible
4	perceptible, but not annoying
3	slightly annoying
2	annoying
1	very annoying

Figure 2: Subjective grading scale

4.2 Experiment

Video sequences were shown on a 4.3" LCD with a resolution of 480 x 272. The observer is positioned approximately 12" to 18" from the display. Observer viewing angle is approximately 10 degrees above normal as shown in Figure 3.

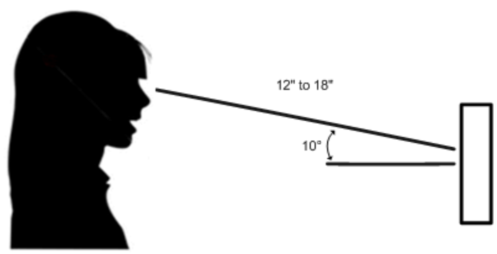

Figure 3: Observer position to LCD

15 observers participated in the test over two separate days. Observers are 18 to 50 years of age and are in good health. Corrective lenses were allowed and used during the test, if the lenses were prescribed.

5. RESULTS

The 640 x 360 video sequences were mapped for spatial perceptual information (SI) and temporal perceptual information (TI) indexes. This will determine high-level characteristics for the video sequence within the frame and motion characteristics between frames. TI and SI calculations are defined in [11]. SI is based on standard deviation over the pixels in each Sobel-filtered frame. The maximum value among all frames is chosen to represent the video sequence. The SI equation is as follows:

$$\boldsymbol{SI = max_{time}\{std_{space}\,[Sobel(F_n)]\}} \tag{1}$$

TI is the motion difference for luminance pixel value in adjacent frames. As with SI, TI is the maximum value among the frames chosen to represent the video sequence. The TI equation is as follows:

$$\boldsymbol{TI = max_{time}\{std_{time}[F_n(i,j) - F_{n-1}(i,j)]\}} \tag{2}$$

T I and SI were calculated for the video sequence with tools made available by [12]. TI and SI graphs are shown in Figure 4 for maximum value, as defined by [11] and Figure **5** for mean value. For short video sequences, TI and SI mean value gives meaningful indication for video temporal expectation. If max value and mean value are fairly close to each other this indicates the sequence temporal changes are fairly consistent throughout the sequence. Scene cuts were removed from the calculation in order to avoid skewing the results. The research team determined results are more meaningful with scene cut data removed. Scene cut removal from calculations is allowed by the ITU standard [11].

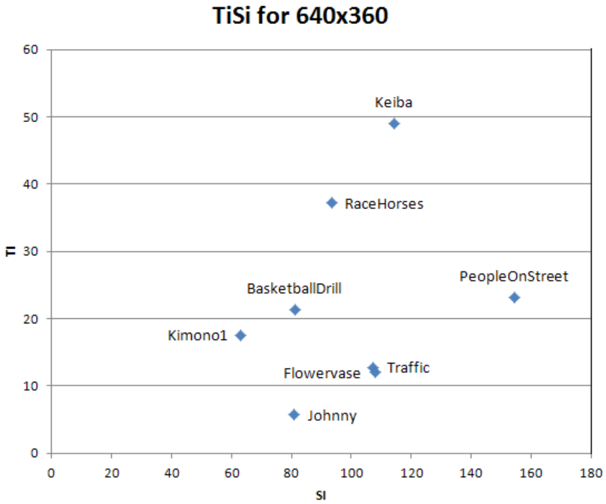

Figure 4: TI SI graph (maximum value)

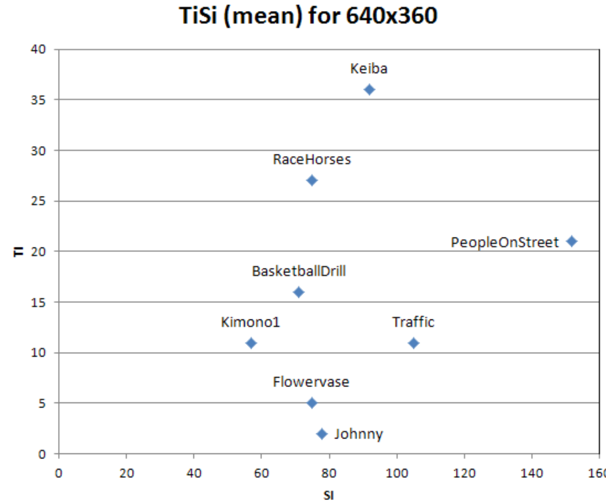

Figure 5: TI SI graph (mean value)

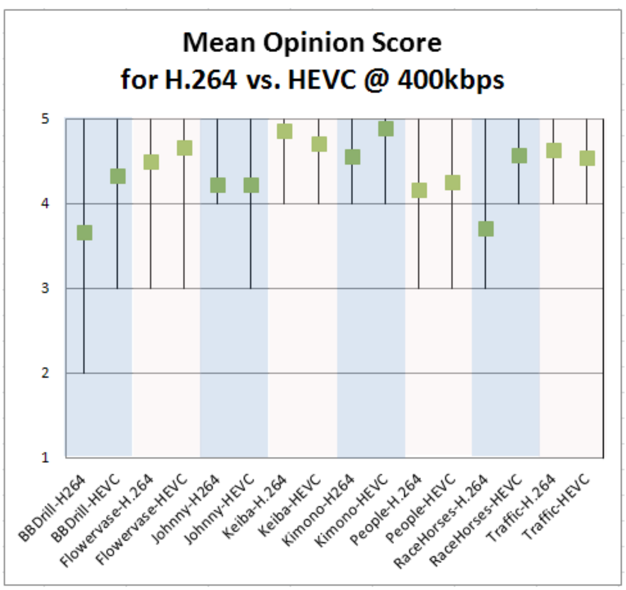

Figure 6: Mean Opinion Score (400kbps)

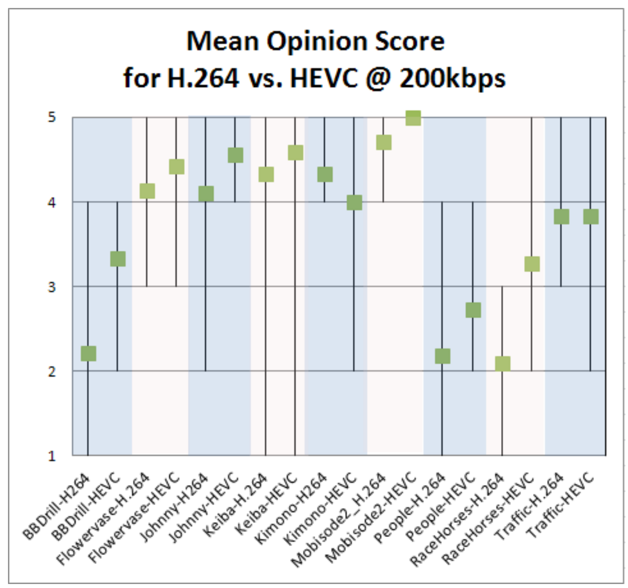

Figure 7: Mean Opinion Score (200kbps)

Subjective feedback varied with the type of content as expected. MOS graphs are shown in Figure 6 and Figure 7 for 400kbps and 200kbps data, respectively. Video sequences with lower TI scores had higher MOS results. This is predominantly due to less need for motion compensation and ability of tested encoders, which leads to higher quality video sequences as noted by observers. For lower bit rate encoded video, the video sequences with higher TI tended to have lower MOS results. However, there are notable exceptions, such as the Keiba video sequence. Which means at lower bit rates, high TI results is a metric that may lead to incorrect conclusions. However, lower TI results is a fairly good indicator of higher MOS scores. The Keiba video sequence is panning with trees in the foreground and rider on a horse, which is also the observer's focal point, moving across the panning screen.

Results show there are predominantly two visual subjective categories. These are "visual obtuse" and "visual acute". "Visual obtuse" is where the observer finds the video sequence busy and difficult to perceive since there are multiple points of action. Essentially, visual obtuse videos have multiple moving regions and no single point of visual attention, such as People on Street, Race Horses, and Traffic video sequences. "Visual acute" is where observers can easily find a single point of attention or there are few points of visual attention. Basically, visual acute videos have single points of visual attention or few moving regions, such as Keiba and Flowervase video sequences. Two classifications were selected for ease of analysis. See Figure 8 and Figure 9 for graphs categorized. Video sequences that have few moving regions have lower TI values which defines the sequences as visual acute. The video sequences that have higher TI values needs a second set of information to determine if video sequences is visual acute or visual obtuse. The additional data needed is the observers' visual attention to moving regions. If there are few points of visual attention, such as the Keiba sequence, then the video is visual acute. If there are many, such as Race Horses, then the video is visual obtuse. "Visual acute" and "visual obtuse" cannot be easily separated in a TiSi Graph as shown in Figure 8. Therefore TiSi graph is not a good representation. The MOS graph yields better results for separating "visual acute" vs. "visual obtuse", however points of attention are not addressed. Figure 10 is a more complete representation of MOS values (shown as point diameter), "points of attention" shown in x-axis, and Ti remains in y axis.

Subjective feedback for 400kbps and category "visual acute" sequences indicate both H.264 and HEVC have a predominant MOS greater than 4, which indicates positive observer viewing experience. Also, the "visual acute" category has MOS difference between HEVC and H.264 of less than 0.25, which indicates little user experience difference. However, for "visual obtuse" category, the MOS delta is greater than 0.5, which indicates a noticeable difference between H.264 and HEVC video sequence's subjective feedback. Also, for the two video sequences, the HEVC observer feedback is around 4.5, which is relatively positive feedback. For H.264 observer feedback is about 3.7, which is approaching marginal feedback. Some MOS separation is starting to occur for "visual obtuse" video sequences at 400 kbps. Nevertheless, the observers list videos encoded by either encoder (H.264 or HEVC) to be subjectively acceptable.

The results at 200 kbps for category "visual acute" indicate video sequences for H.264 and HEVC have MOS at 3.8 and higher. The MOS drop is expected when compared with 400kbps. The MOS gap between H.264 and HEVC video sequences has increased to 0.5. This indicates the observer's experience is starting to lean towards favoring HEVC, however, observers still found either H.264 or HEVC to be mainly acceptable. "Visual Obtuse" sequences at 200kbps had MOS scores degrade severely for both H.264 and HEVC. This indicates the encoder struggled to display subjectively acceptable video. "Visual obtuse" category video sequences with higher TI tended to see a wider gap in MOS results. This implies video motion causes subjective scores to degrade faster for H.264 encoded sequences.

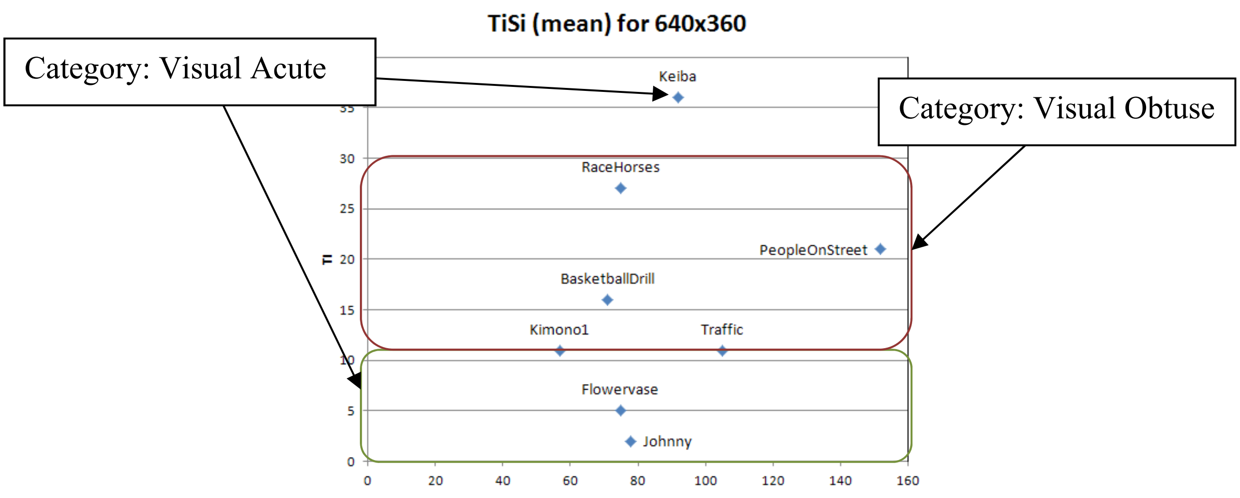

Figure 8: Analyzed TI SI graph

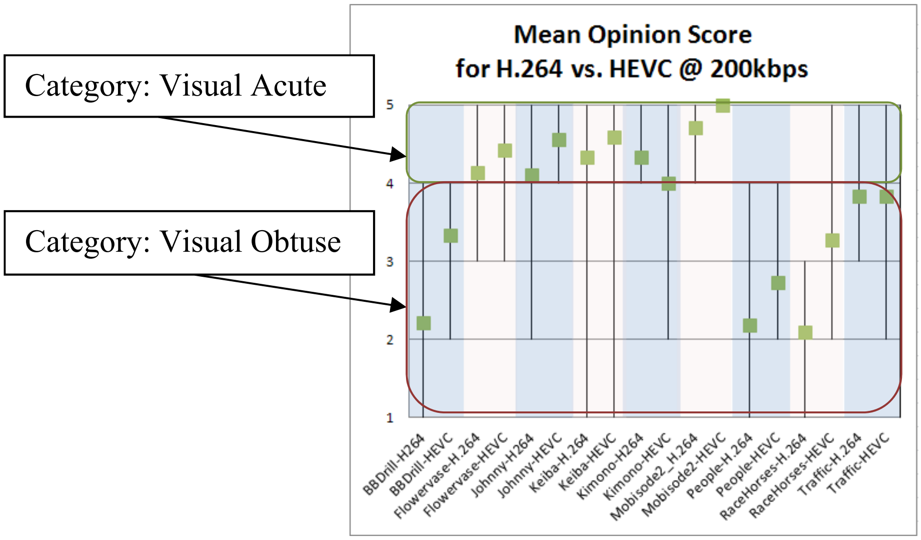

Figure 9: Analyzed MOS graph

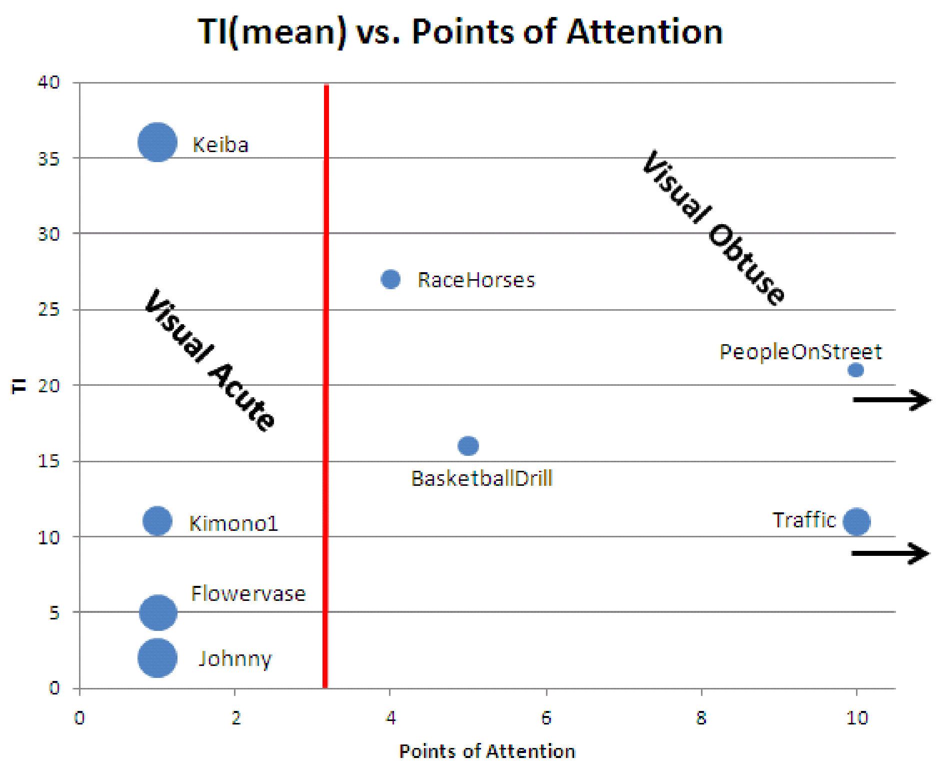

Figure 10 : TI(mean) vs. Points of Attention. The size of the data point represents MOS value for HEVC at 200kbps. Note: "PeopleOnStreet" and "Traffic" are to the right of the graph, as denoted by the arrows.

6. CONCLUSION

For mobile resolution, such as 640 x 360, subjective feedback shows both encoding methods are adequate at 400kbps constant bit rate. A consumer experience gap was observed for 200 kbps constant bit rate. Significantly less H.264 subjective quality is noticed with video sequences that have multiple objects moving and no single point of visual attraction. Video sequences with single points of visual attraction or few moving objects tended to have H.264 encoded video on par with HEVC encoded video.

REFERENCES

[1] Vittorio Baroncini, Jens-Rainer Ohm, and Gary Sullivan, "Report of Subjective Test Results of Responses to the Joint Call for Proposals (CfP) on Video Coding Technology for High Efficiency Video Coding (HEVC)", Joint Collaborative Team on Video Coding (JCT-VC) of ITU-T SG16 WP3 and ISO/IEC JTC1/SC29/WG11, JCTVC-A204, pp:1-33, April2010

[2] ITU-R, "Methodology for the subjective assessment of the quality of television pictures", Recommendation ITU-R BT.500-13, ITU-R BT.500-13, pp:1-46, January2012

[3] TK Tan, A. Fujibayashi, Y. Suzuki, and J.Takiue, "Objective and subjective evaluation of HM5.0.", Joint Collaborative Team on Video Coding (JCT-VC) of ITU-T SG16 WP3 and ISO/IEC JTC1/SC29/WG11, JCTVC-H0116, 1-24, February2012

[4] Jens-Rainer Ohm, Gary Sullivan, Frank Bossen, Thomas Wiegand, Vittorio Baroncini, Mathias Wien, and Jizheng Xu, "JCT-VC AHG report: HM subjective quality investigation (AHG22)", Joint Collaborative Team on Video Coding (JCT-VC) of ITU-T SG16 WP3 and ISO/IEC JTC1/SC29/WG11, JCTVC-H0022r1, 1-3, February2012

[5] Michael Horowitz, Faouzi Kossentini, Nader Mahdi, Shilin Xu, Hsan Guermazi, Hassene Tmar, Bin Li, Gary Sullivan, Jizheng Xu, "Informal Subjective Quality Comparison of Video Compression Performance of the HEVC and H.264 / MPEG-4 A VC Standards for Low-Delay Applications", SPIE Applications of Digital Image Processing XXXV, Vol:8499, 84990W, pp:1-6, August2012

[6] Philippe Hanhart, Martin Rerabek, Francesca De Simone, and Touradj Ebrahimi, "Subjective Quality Evaluation of the Upcoming HEVC Video Compression Standard", Retrieved 2012Oct10 from http://www.itu.int/ITU-T/studygroups/com16/jct-vc/, pp:1-13, August2012

[7] Brightcove, "Encoding for Mobile Delivery", Retrieved 2012July08 from http://support.brightcove.com/en/docs/encoding-mobile-delivery

[8] Apple, "iOS Developer Library - Preparing Media for Delivery to iOS-Based Devices", Retrieved 2012July08 from https://developer.apple.com/library/ios/#documentation/NetworkingInternet/Conceptual/StreamingMediaGuide/UsingHTTPLiveStreaming/UsingHTTPLiveStreaming.html#//apple_ref/doc/uid/TP40008332-CH102-SW8

[9] Fraunhofer Institut Nachrichtentechnik Heinrich-Hertz Institut, "H.264/AVC Software Coordination Version: JM18.3", Retrieved 2012May09 from http://iphome.hhi.de/suehring/tml/

[10] Fraunhofer Institut Nachrichtentechnik Heinrich-Hertz Institut, "High Efficiency Video Coding (HEVC) version 6.0", Retrieved 2012May06 from http://hevc.hhi.fraunhofer.de/

[11] ITU-T, "Subjective video quality assessment methods for multimedia applications", Recommendation ITU-T P.910, ITU-T P.910, pp:1-42