QoE Evaluation of Multimedia Transmission over Wireless Networks

Xiaojing Li, Kai Dong, Wei Song, and Bradford G. Nickerson
Faculty of Computer Science
University of New Brunswick, Fredericton, Canada
Emails: {xiaojing.li, kdong, wsong, bgn}@unb.ca

Abstract—Many simulation tools (e.g., NS-2) can examine the quality of service (QoS) of networks but cannot demonstrate the visual and auditory effects of wireless transmission on multimedia quality. This paper presents a wireless multimedia simulator (WMS), which uses a compact graphical user interface to present the real-time packet delay with the playback of streaming media over a wireless channel based on classic radio channel models and IEEE 802.11 medium access control. By using captured packets and reproduced traces, WMS can demonstrate the visual and auditory effects of fading errors, packet delay and loss. Leveraging the real-time playback function, WMS enables quality of experience (QoE) evaluation of multimedia transmission in a controlled wireless environment.

We carried out QoE evaluation with 30 participants for 104 test cases comprising 2 videos and 2 audio clips produced by WMS. The valuable test results enable us to quantify the relationship of QoE in terms of mean opinion score (MOS) with network traffic load and QoS metrics such as bit error probability (BEP). We find the subjective QoE is sensitive to media content although consistent with objective QoS metrics. Statistical difference-of-means tests show the video with slower motion and fewer colours is likely to offer better delay tolerance, and audio is less sensitive to bit errors while video is more resistant to network congestion.

Index Terms—Graphical network simulator, multimedia transmission, quality of service (QoS), quality of experience (QoE).

I. INTRODUCTION AND RELATED WORK

With the explosive growth of wireless networks in recent years, multimedia services are one of the most popular and dominant applications on mobile devices. Due to the limited and varying bandwidth of wireless networks, the quality of multimedia delivery cannot be as good or stable as in wired networks. To examine the effects of wireless networks on multimedia service quality, many existing simulators focus on objective quality of service (QoS) metrics, such as bit error probability (BEP), packet delay, and packet loss. It is the quality of experience (QoE), however, that ultimately determines the user-perceived service quality. As defined in [1] by the Telecommunication Standardization Sector of the International Telecommunication Union (ITU-T), QoE is "the overall acceptability of an application or service, as perceived subjectively by the end-user."

EVOM [2] assesses the perceived quality of transmitted voice over a wireless mobile ad-hoc network. It employs NS-2 for network simulation with different velocities, node densities, background traffic, and transmission ranges. EVOM can assess the quality of degraded voice generated with codecs G.711 and G.729, but it does not support video quality evaluation. EvalVid [3] is a video quality evaluation framework for a real or simulated network. It assesses the video quality of MPEG-4 streaming by computing the peak signal-to-noise ratio (PSNR) of degraded video and mapping it to mean opinion score (MOS). EvalVid is extended in [4] by replacing PSNR with the fraction of decodable frames. EvalVid also integrates NS-2 for network simulation. Video Tester [5] is another video quality evaluation simulator that aims at providing multiple quality metrics such as PSNR and PSNR-based MOS estimates. Video Tester uses a third-party video codec library and does not include a network simulation module.

The opinion scores or ratings used in the above simulators to assess multimedia quality are actually computed from some mapping models of unjustified accuracy. These simulators did not conduct real subjective user tests, which can provide more accurate and complete results. This is mainly because subjective QoE experiments are more challenging and expensive. First of all, a specialized simulator is required to support real-time playback of multimedia content and demonstrate the visual and auditory effects of wireless transmission. Furthermore, the QoE test cases should be designed appropriately so that the number of tests is minimized while the distortions due to different factors such as BEP, packet loss rate, and packet delay can still be clearly exposed. The QoE evaluation has to follow strict experimental guidelines such as [6] in a controlled environment without disturbance. Last but not the least, a reasonable number of participants are needed to produce sufficient test data, which should be analyzed rigorously to derive in-depth insights.

Based on these observations, we developed a novel wireless multimedia simulator (WMS) to facilitate multimedia quality assessment. WMS implements a typical wireless physical channel and the IEEE 802.11 contention-based medium access control (MAC). Objective metrics, including BEP, packet loss rate, and packet delay are dynamically computed to provide accurate wireless transmission status. The multimedia content can be played back in real time by the simulator to directly show the visual and auditory effects of the wireless transmission. We conducted formal subjective assessments of the wireless transmission impairment with 30 participants for 104 test cases comprising 2 videos and 2 audio clips produced by WMS. These extensive and valuable assessment results were analyzed and presented in various formats including radar charts that quantify the relationship of a subjective QoE measure (i.e.
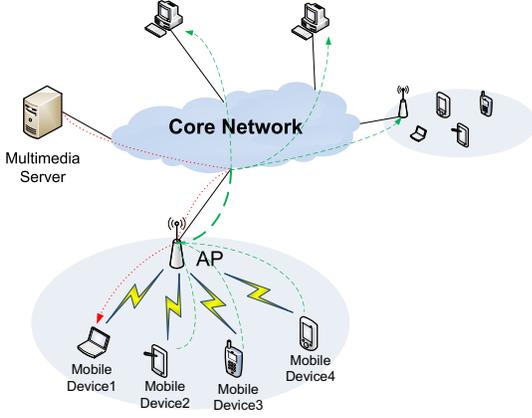
Fig. 1. Simulation scenario of multimedia delivery over wireless networks.

MOS) with objective QoS metrics (i.e. BEP and traffic load). To our knowledge, WMS is the first tool supporting subjective QoE evaluation and assisting wireless network configuration to deliver satisfactory multimedia services.

II. SYSTEM MODEL

Fig. 1 shows our simulation scenario, in which `Mobile Device1` is receiving and playing in real time streaming audio or video from a multimedia server across the core network. The other mobiles are transmitting saturated background data traffic through the access point (AP). We focus on the wireless link and assume an ideal wired connection between the multimedia server and the AP.

A. Channel Modelling

WMS incorporates three classic radio channel models, namely, the free space model, log-normal shadowing model, and Rayleigh fading model [7]. The path loss PL at a transmitter-receiver separation distance d is given by

$$PL(d) = \overline{PL}(d_0) + 10n \log \left(\frac{d}{d_0} \right) + X_\sigma \tag{1}$$

where $\overline{PL}(d_0)$ is the path loss measured at a reference distance d_0 close to the transmitter, n is the path-loss exponent depending on the environment, and X_σ is a zero-mean Gaussian random variable with standard deviation σ. Rayleigh fading is also considered to model the small-scale fading effect.

We consider M-ary quadrature amplitude modulation (M-QAM) as the modulation scheme, where $M \leq 128$ since modulations beyond 128-QAM can be complex and challenging in practice. Given the average ratio ($\overline{\gamma}$) of the signal energy per symbol to noise power spectral density (E_s/N_0), the average BEP of Rayleigh fading channel is given by [8]

$$BEP \approx \frac{1}{2} c_{Mi} \left(1 - \sqrt{\frac{d_{Mi}^2 \overline{\gamma}}{1 + d_{Mi}^2 \overline{\gamma}}} \right) \tag{2}$$

where M_i is the modulation constellation size and $M_i = 2^i$

$(i = 1, 2, ..., 7)$, and c_{Mi} and d_{Mi} are given by

$$c_{Mi} = \begin{cases} 1, & M_i = 2 \\ 2\dfrac{1 - 1/\sqrt{M_i}}{\log_2 \sqrt{M_i}}, & M_i \geq 4 \end{cases} \tag{3}$$

$$d_{Mi} = \begin{cases} 1, & M_i = 2 \\ \sqrt{\dfrac{3}{2(M_i - 1)}}, & M_i \geq 4. \end{cases} \tag{4}$$

For the MAC layer, our simulator uses an event-driven implementation of the basic mode of distributed coordination function (DCF) of IEEE 802.11. DCF is based on carrier sense multiple access with collision avoidance (CSMA/CA), which manages to avoid collision by sensing the physical channel, using random backing off timer, and distinguishing inter-frame space in different situations.

B. Video and Audio Traffic Specification

WMS currently supports the widely used H.264, advanced audio coding (AAC), and MPEG audio layer III (MP3) for video/audio coding, and MPEG-2 transport stream (TS) as the container format. Multimedia streaming is based on `MPEG-2 TS/RTP/UDP/IP`, in which the real-time transport protocol (RTP) runs over the user datagram protocol (UDP) to deliver media data over Internet protocol (IP)-based wireless networks. The media steaming and playback are implemented with the open source libraries VLC [9] and VLCJ [10].

III. SIMULATOR DESIGN AND IMPLEMENTATION

A. Simulator Framework

To facilitate subjective QoE assessment, we develop a specialized wireless multimedia simulator (WMS). WMS can incorporate bit errors, packet loss and packet delay into the original video/audio streaming packets, so that the reproduced packet traces are played back to demonstrate the transmission impairments. Meanwhile, a dynamic plot shows the objective QoS of packet delay and a built-in VLC player shows the subjective effects of these objective metrics in real time. Fig. 2 shows the simulation process for a multimedia streaming flow.

A VLCJ server is employed to encode and stream local raw media files [10]. Before the media content is sent to the network interface, our simulator uses a Java library Jpcap to capture the packets at the link layer and writes the packet trace into a local .jpcap file [11]. After that, the simulator uses self-developed network simulation functions to process the .jpcap file. To introduce transmission impairment caused by bit errors, the bit error generator examines the .jpcap file bit by bit and flips each bit according to BEP. BEP is obtained by the channel simulator for the physical layer and depends on various channel parameters. Thus, we can reproduce a new .jpcap packet trace with simulated bit errors.

In our simulator, we focus on the contention-based MAC of IEEE 802.11, in which the contentions and collisions among mobile stations can result in packet loss and delay. By simulating the channel access procedure, we obtain the drop flag, which indicates whether a packet is lost, and the packet leaving time, which is the moment that a packet is successfully
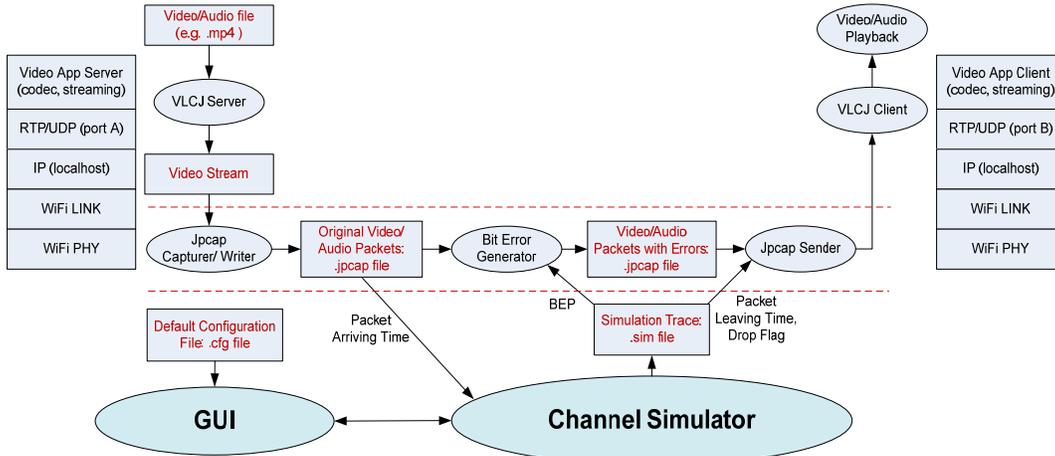
Fig. 2. Processing flow of multimedia packets in the WMS simulator.

transmitted. In the graphical user interface (GUI), the real-time packet delay is shown in a dynamic plot by JFreeChart [12], a Java library for drawing various charts.

To provide a subjective perception of the effect of objective metrics on a multimedia stream, we use Jpcap to retrieve packets from the degraded .jpcap file and deliver them one by one to resume the multimedia stream. That is, the packets with possible bit errors are passed to the VLCJ client according to the drop flag and packet leaving time obtained from the channel simulator. The client then starts to play back degraded packets it has received, while the following packets are processed in the channel simulator module. This design enables users to examine the degraded multimedia in real time rather than wait until the network simulation process ends. As such, the visual/auditory effects of wireless transmission impairments can be properly presented together with the objective QoS metrics. This property enables operators to view or listen to the degraded multimedia and follow the packet delay situations at the same time, which provides a good way to analyze the cause of multimedia quality degradation.

B. Simulator Graphical User Interface

Fig. 3 shows the GUI interface where users interact with the simulator. The top left is a section for specifying the location of the files to be utilized and generated. On the bottom left is a section for specifying the parameters of the MAC layer and the physical layer. On the top right is an embedded VLC media player for the streaming video/audio playback. On the left side of the player, a small panel displays the attributes of original video/audio. In the middle of the right panel is a dynamic plot showing the real-time delay of the packets when they are being transmitted and played back. On the bottom right are two small panels. One shows the physical layer status including actual data transmission rate, the modulation scheme, and BEP. The other displays current simulation progress.

IV. QoE EXPERIMENT RESULTS

Fading errors and traffic load are two critical factors which influence the received multimedia quality. Due to the contention-based channel access, a higher traffic load may involve a larger transmission overhead and collisions, which result in packet delay and loss. This section demonstrates their effects by changing the path loss exponent (n), transmitter-receiver distance (d), and number of stations. The other physical and MAC parameters are given in Table I.

TABLE I
DEFAULT SYSTEM PARAMETERS.

Parameter	Value
Background flow packet length	6000 bits
Reference distance	1 m
Channel frequency	2.4 GHz
Transmitter antenna gain	1
Receiver antenna gain	1
Transmit power	30 dBm
Shadowing standard deviation	0 dB
Average noise power	-110 dBm

A. Effect of Fading Errors, Packet Loss, and Delay

To demonstrate the effect of fading errors, we use parameters path loss exponent (n) and distance (d) to simulate channel conditions giving different BEPs using equation (2). With 2 mobile stations, a data rate of 11 Mbit/s, and BPSK for modulation, Fig. 4 shows how a high BEP results in frame distortion and incorrectly recovered video pictures, which may seriously affect the perceived quality.

By simulating a large number of stations within the coverage of an AP, we can evaluate the effects of packet loss and delay. We find that packet loss has a similar effect to bit errors on video playback, except that the picture distortion may appear in a blurred area rather than randomly across the picture. This is because the lost packets may originate from a coded block. Likewise, we can observe the effect of packet delay. When there is a large number of stations (e.g., 30), evident interruptions occur because the packets of the next frame to play cannot arrive in time. Figures illustrating these effects on video playback can be found in [13].

B. Effect on Quality of Experience

The results in Section IV-A validate the design of our simulator, which effectively demonstrates the visual and auditory
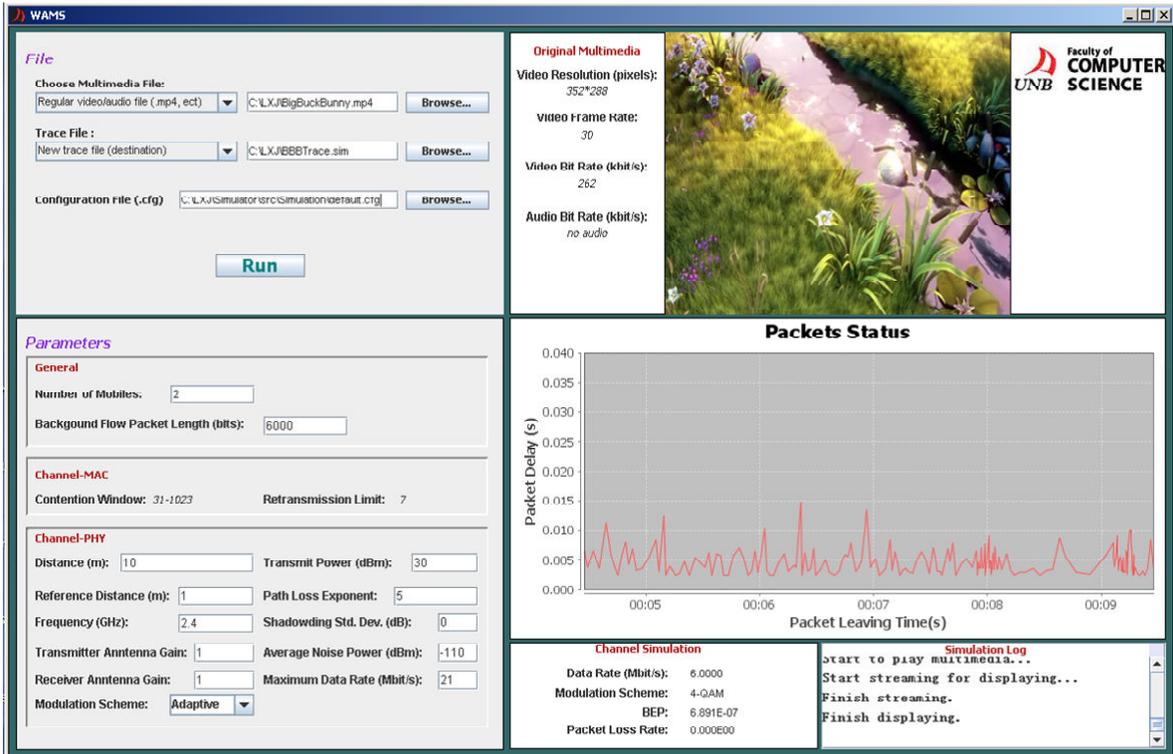
Fig. 3.   Simulator GUI.

Fig. 4.   Simulated "News" video transmission illustrating the effect of fading errors with two different channel conditions, (a) BEP=1.1×10^{-8} ($n = 2$, $d = 30$ m) vs. (b) BEP=1.3×10^{-5} ($n = 3$, $d = 100$ m).

effects as well as the real-time display of the objective QoS metrics. This enabled us to conduct subjective experiments to evaluate the effects of wireless transmission (packet delay, packet loss, and bit errors) on ultimate QoE. The experiments consisted of 104 test cases which are produced by WMS based on 2 videos without audio track ("News", "Mobile") and 2 audio clips ("Speech", "Piano"). The videos are coded by H.264, while the audio clips are coded by MP3. Each of the 4 media files generates 25 test cases with simulated transmission impairments and one original stream without any simulated delay or bit errors. Following the guidelines in [6], we performed the tests in the Usability Lab of the Faculty of Computer Science at UNB, which is designed to carry out usability tests without disturbance. The tests lasted for 52 days with 30 non-expert participants. The participants assessed a random sequence of these videos and audio clips and rated them based on a five-point scale [6], in which 5 stands

for *Imperceptible*, 4 for *Perceptible, but not annoying*, 3 for *Slightly annoying*, 2 for *Annoying*, and 1 for *Very annoying*.

The average ratings of 30 participants for each case (except the four original ones) are presented in Figs. 5-6. In Fig. 5 we can see that the MOS of "Mobile" is much less than that of "News", indicating that "News" is more tolerant to bad channel conditions. Comparing Fig. 5 and Fig. 6 we can observe that the MOS decreases with the increase of BEP or number of stations. This observation confirms that bit errors and the traffic load (corresponding to packet delay and loss) indeed negatively affect user-perceived video/auditory quality.

We also show the results in a more intuitive way by the radar charts in Figs. 7-8. The length axis represents the number of stations, while the angle axis represents BEP. The grey levels of the shaded areas represent the MOS. For comparison purposes, we show the MOS of different videos or audio clips in the upper half and lower half of a radar chart.

Comparing the top and bottom of Fig. 7 for two videos, we observe that the area of "News" in each level of MOS is much larger than that of "Mobile", which confirms that "News" is more tolerant to bad channel conditions. This is because "News" contains slower motion and far fewer colours than "Mobile" and results in less data for transmission. Examining the trace files, we find that the original "News" video lasts for 10s and produces 259 streaming packets in total. In contrast, "Mobile" is of the same duration but produces 1086 packets. The videos of "News" and "Mobile" generate packets at an average rate of 25.9 and 83.2 packets/s, respectively. That is, "News" produces packets at a much lower rate than "Mobile" and needs less bandwidth to achieve the same quality.

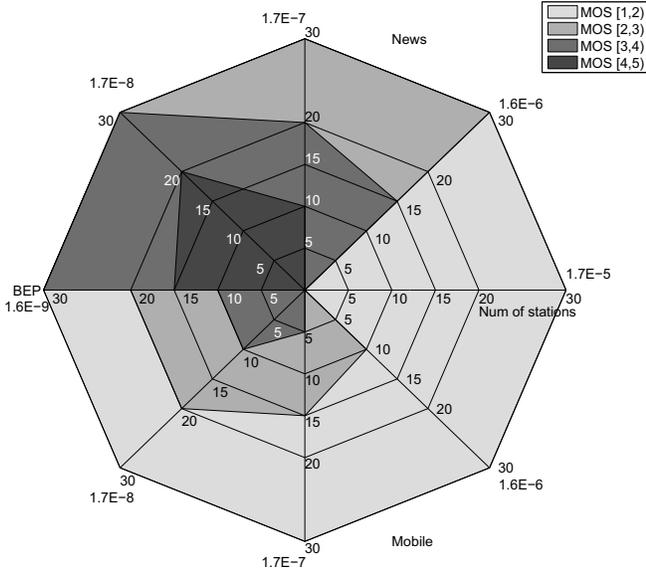Comparing the top and bottom of Fig. 8 for the audio cases,

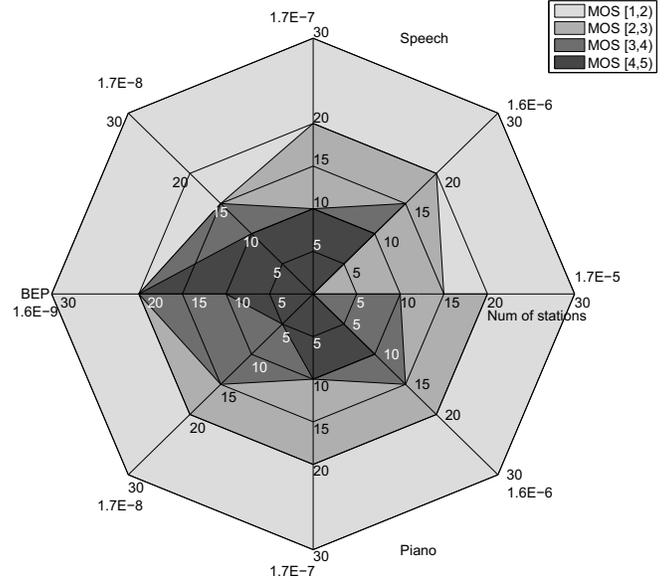Fig. 7.   Radar chart of MOS for video clips (top: "News"; bottom: "Mobile").

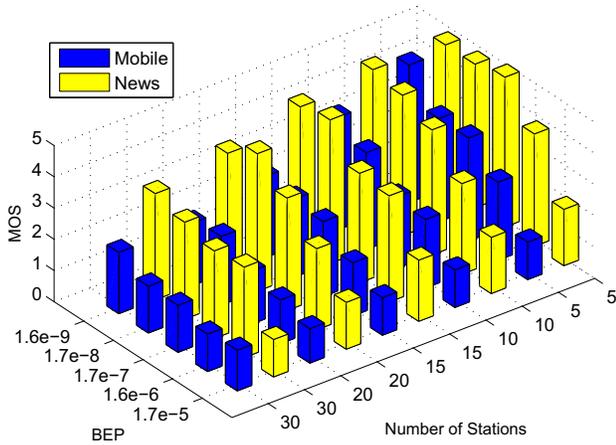Fig. 8.   Radar chart of MOS for audio clips (top: "Speech"; bottom: "Piano").
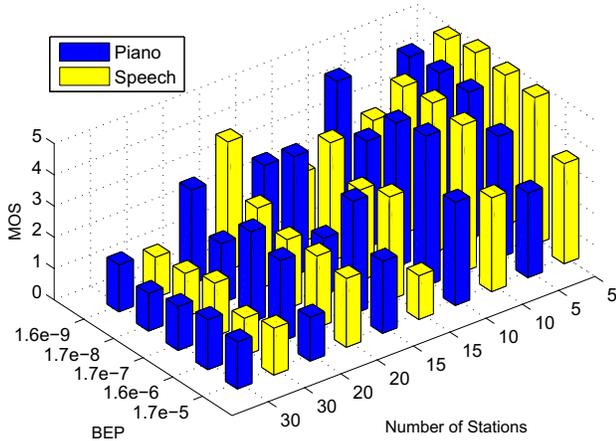

Fig. 5.   MOS of video.

Fig. 6.   MOS of audio.

we find the area of "Speech" in each level of MOS is much larger than that of "Piano", except for the MOS level [3,4], in which both have the same area. This is consistent to our intuition that "Speech" is more tolerant to network distortion than "Piano" since people have a higher quality expectation and sensitivity to musical content compared to human voice.

Comparing Fig. 7 and Fig. 8 for video and audio, we see that MOS along BEP degrades more severely for video than for audio with a small number of stations. For example, when the number of stations is 10 and BEP increases from 1.6×10^{-9} to 1.7×10^{-5}, the MOS of "News" drops by 3 levels from [4,5] to [1,2]. The MOS of both "Mobile" and "Speech" drops by 2 levels from [3,4] to [1,2] and from [4,5] to [2,3], respectively. The MOS of "Piano" drops by 1 level from [4,5] to [3,4]. This indicates that video is more sensitive to random loss caused by bit errors than audio when the traffic load is not heavy.

On the other hand, MOS along the axis of number of stations degrades more severely for audio when BEP is not very large. For example, when BEP is 1.7×10^{-8} and the number of stations increases from 5 to 20, the MOS of "News" does not change, the MOS of "Mobile" drops by 1 level from [3,4] to [2,3]. In contrast, the MOS of "Speech" drops by 2 levels from [4,5] to [2,3], while the MOS of "Piano" drops by 3 levels from [4,5] to [1,2]. This observation indicates that audio is more sensitive to packet loss and delay (caused by network congestion with a large number of stations) than video when bit errors are not severe.

As people are likely to have different subjective opinions to the perceived multimedia quality, the scores for each test case vary among the 30 participants. Based on the mean and variance of each test case, we used a statistical difference-of-means test [14] to analyze the MOS difference between simulated multimedia traces and the corresponding originals. We construct a hypothesis H_0 that the MOS of a simulated case is equal to the MOS of the corresponding original. The alternative hypothesis H_1 is that the MOS of the simulated

TABLE II
STATISTICAL ANALYSIS RESULTS FOR MOS OF EACH TEST CASE.

Number of Stations	BEP	MOS Statistical Validation			
		News	Mobile	Speech	Piano
5	1.6×10^{-9}	▲	▲	▲	▲
5	1.7×10^{-8}	▲	<	▲	▲
5	1.7×10^{-7}	▲	<	▲	▲
5	1.6×10^{-6}	<	<	▲	<
5	1.7×10^{-5}	<	<	<	<
10	1.6×10^{-9}	▲	<	<	▲
10	1.7×10^{-8}	▲	<	▲	<
10	1.7×10^{-7}	<	<	▲	▲
10	1.6×10^{-6}	<	<	▲	▲
10	1.7×10^{-5}	<	<	<	<
15	1.6×10^{-9}	▲	<	<	<
15	1.7×10^{-8}	▲	<	<	<
15	1.7×10^{-7}	<	<	<	<
15	1.6×10^{-6}	<	<	<	<
15	1.7×10^{-5}	<	<	<	<

case is lower than that of the original. Assuming that the test samples follow a t-distribution, we can accept either H_0 or H_1 as the relationship between the MOS of a simulated test case and the original, with a certain confidence level.

Table II shows the results of the statistical analysis at the 0.01 level of significance. There are 21 cases that have the same perceived quality (denoted by "▲") as the original when the number of stations varies from 5 to 15. The other cases having a perceived quality worse than that of the original are denoted by "<". Comparing "News" and "Mobile", we can see that more cases of "News" (7 vs. 1) are perceived to be of equal quality to the original file, which confirms that the former is more resistant to bad channels. Comparing "News", "Speech", and "Piano", when the number of stations is 15, "News" still has 2 cases with an equal quality to the original file, while neither "Speech" nor "Piano" has such an equal case. We thus see that "News" is more resistant to packet loss and delay caused by network congestion when bit errors are not severe. On the other hand, when the number of stations is 10 and BEP is larger than 1.7×10^{-8}, we still find 2 cases each in "Speech" and "Piano" that have an equal quality to the original file, but none is found in "News". This observations shows that "Speech" and "Piano" are more resistant to random loss caused by bit errors under a low traffic load.

We also observe some abnormal cases in the radar charts and Table II, such as with "Speech" (10, 1.6×10^{-9}) (the two-tuple denotes the number of stations and BEP, respectively), "Piano" (10, 1.7×10^{-8}), and "Piano" (5, 1.6×10^{-6}). We examined the test audio files and found that there is some noise at the end of the audio, due to the encoding loss or mechanical error, which is irrelevant to the simulated networks. However, when the network condition is good enough, such random noise will make a difference on the final MOS. To address such a problem, it would be better to consider including a brief silence episode (less than 1s) at the beginning and end of the audio in future experiments.

V. CONCLUDING REMARKS

In this work, we carried out accurate QoE evaluation for the quality of multimedia transmission over wireless networks.

To support subjective user tests, we designed a specialized wireless multimedia simulator (WMS), which displays the real-time packet delay in a dynamic plot, and plays back the video/audio content showing the visual/auditory transmission impairments. WMS can effectively demonstrate the effects of bit errors, packet delay, and packet loss on multimedia quality in a simulated wireless networking environment.

Following strict experiment guidelines, we conducted subjective QoE tests with 30 participants for 104 test cases. As QoE tests are time-consuming, we carefully selected the test cases so that the effect of each particular factor can be isolated and clearly exposed. Based on the test results, we can quantify the relationship of subjective QoE in terms of MOS with objective QoS metrics such as BEP and traffic load. Our experiment results further show that the video with slower motion and fewer colours has a better delay tolerance than those with fast movement and colourful scenes. We also find that video is more sensitive to random bit errors, while audio is more sensitive to packet loss and delay caused by network congestion. Our statistical test with the QoE results also verifies these conclusions.

REFERENCES

[1] "Definition of quality of experience (QoE)," ITU-T, ITU-R P.10/G.100 Amendment 1, Jan. 2007.
[2] S. Jelassi and H. Youssef, "EVOM: A software based platform for voice transmission and quality assessment over wireless ad-hoc networks," in *Proc. 10th ACM Symposium on Modeling, Analysis, and Simulation of Wireless and Mobile Systems*. ACM, 2007, pp. 169–172.
[3] J. Klaue, B. Rathke, and A. Wolisz, "EvalVid - A framework for video transmission and quality evaluation," in *Computer Performance Evaluation. Modelling Techniques and Tools*. Springer, 2003, pp. 255–272.
[4] C.-H. Ke, C.-K. Shieh, W.-S. Hwang, and A. Ziviani, "An evaluation framework for more realistic simulations of MPEG video transmission." *Journal of Information Science and Engineering*, vol. 24, no. 2, pp. 425–440, 2008.
[5] I. Ucar, J. Navarro-Ortiz, P. Ameigeiras, and J. M. Lopez-Soler, "Video Tester - A multiple-metric framework for video quality assessment over IP networks," in *Proc. IEEE International Symposium on Broadband Multimedia Systems and Broadcasting (BMSB)*, 2012.
[6] "Methodology for the subjective assessment of the quality of television pictures," ITU-T, ITU-R BT.500-13, Jan. 2012.
[7] T. S. Rappaport, *Wireless Communications: Principles and Practice*, 2nd ed. Prentice Hall, 2002.
[8] A. B. Sediq and H. Yanikomeroglu, "Performance analysis of selection combining of signals with different modulation levels in cooperative communications," *IEEE Trans. Veh. Technol.*, vol. 60, no. 4, pp. 1880–1887, May 2011.
[9] VideoLAN, "VLC," http://www.videolan.org/vlc/index.html, accessed Dec 4, 2012.
[10] Caprica Software Limited, "VLCJ," http://www.capricasoftware.co.uk /vlcj/, accessed Dec 4, 2012.
[11] K. Fujii, "Jpcap," http://netresearch.ics.uci.edu/kfujii/Jpcap/doc/, accessed Aug. 1, 2012.
[12] Object Refinery Limited, "JFreeChart," http://www.jfree.org, accessed Dec 4, 2012.
[13] X. Li, "Simulator for adaptive multimedia transmission over wireless networks," Master of Computer Science Thesis, University of New Brunswick, Fredericton, Canada, Jan. 2013.
[14] R. E. Walpole and R. H. Myers, *Probability and Statistics for Engineers and Scientists*. The Macmillan Company, 1972.